

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
25 April 2002 (25.04.2002)

PCT

(10) International Publication Number  
WO 02/33099 A2(51) International Patent Classification<sup>7</sup>: C12N 15/54, 9/12, A01K 67/027, C07K 16/40, C12Q 1/68, 1/48, G01N 33/50, A61K 38/45

(21) International Application Number: PCT/US01/47728

(22) International Filing Date: 20 October 2001 (20.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/242,410	20 October 2000 (20.10.2000)	US
60/244,068	27 October 2000 (27.10.2000)	US
60/245,708	3 November 2000 (03.11.2000)	US
60/247,672	9 November 2000 (09.11.2000)	US
60/249,565	16 November 2000 (16.11.2000)	US
60/252,730	22 November 2000 (22.11.2000)	US
60/250,807	1 December 2000 (01.12.2000)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GURURAJAN, Rajagopal [US/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). WALIA, Narinder, K. [US/US]; 890 David Street, #205, San Leandro, CA 94577 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). ARVIZU, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). RAMKUMAR, Jaya [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). DING, Li [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). NGUYEN, Dannie, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). YUE,

Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). TRI-BOULEY, Catherine, M. [FR/US]; 1121 Tennessee, #5, San Francisco, CA 94107 (US). LAL, Preeti, G. [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). RECIPON, Shirley, A. [US/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062 (US). SWARNAKER, Anita [CA/US]; 8 Locksley Avenue, #5D, San Francisco, CA 94122 (US). THANGAVELU, Kavitha [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94304 (US). KHAN, Farrah, A. [IN/US]; 3617 Central Road, #102, Glenview, IL 60025 (US). ISON, Craig, H. [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

## Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HUMAN KINASES

(57) Abstract: The invention provides human human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PKIN.

WO 02/33099 A2

## HUMAN KINASES

## TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human kinases and to the  
5 use of these sequences in the diagnosis, treatment, and prevention of cancer, immune disorders,  
disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the  
assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid  
sequences of human kinases.

10

## BACKGROUND OF THE INVENTION

Kinases comprise the largest known enzyme superfamily and vary widely in their target  
molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to  
a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most  
kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of  
15 molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are  
phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the  
acceptor molecule, causing internal conformational changes and potentially influencing intermolecular  
contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in  
eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular  
20 signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated  
proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second  
messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens,  
that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such  
25 as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular  
environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells  
has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle  
have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked  
to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

30

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs),  
phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs),  
phosphorylates serine and threonine residues. Some PTKs and STKs possess structural  
characteristics of both families and have dual specificity for both tyrosine and serine/threonine

residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in 15 Hardie, G. and Hanks, S. (1995) The Protein Kinase Facts Book, Vol I p.p. 17-20 Academic Press, San Diego, CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% 20 (PROSITE: PDOC00100, November 1995).

#### Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which 25 causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form 30 signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK

activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and Tonks, N. K. (1992) *Annu. Rev. Cell Biol.* 8:463-93). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

5 **Protein Serine/Threonine Kinases**

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors.

10 Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down- regulation.

15 Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases

20 involved in apoptosis.

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP-ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York, NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein

kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism, and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al., (1995) *J. Biol. Chem.* 5 270:14875-14883.

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al., (1998) *J. Biol. Chem.* 273:1357-1364.) The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of 10 casein kinase I from a variety of sources. There are at least five mammalian isoforms,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ . Fish et al., identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by 15 deletion of the yeast CKI locus, HRR250 (Fish et al, *supra*.)

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the 20 mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state. Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al., have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag 25 and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al., (2000) *Science* 288:483-491.)

#### Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle 30 contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by

other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO Journal 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the 5 synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both 10 axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

#### Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP) which mediate signal transduction from the cell 15 surface to the nucleus via phosphorylation cascades are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and Weinberg, R.A. (1993) Nature 365:781-783). MAP kinase signaling pathways are present in mammalian cells as well as in yeast. The extracellular stimuli which activate MAP kinase pathways include epidermal growth 20 factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS), and pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.

#### Cyclin-Dependent Protein Kinases

25 The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition 30 to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation

of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M., et al. (1998) EMBO J. 17:470-481). The NIM-related kinases also include NIK1 histidine kinases, which function in signal transmission (Yamada-Okabe, T. et al. (1999) J. Bacteriol. 181:7243-7247).

#### Checkpoint and Cell Cycle Kinases

5 In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant  
10 cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition. (Sanchez, Y. et al. (1997) Science 277:1497-1501.) Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is  
15 to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis. (Peng, C-Y et al. (1997) Science 277:1501- 1505.) Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A similar deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

20 Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-8). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor  
25 tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

The RET (rearranged during transfection) proto-oncogene encodes a tyrosine kinase receptor involved in both multiple endocrine neoplasia type 2, an inherited cancer syndrome, and Hirschsprung disease, a developmental defect of enteric neurons. RET and its functional ligand, glial cell line-derived neurotrophic factor, play key roles in the development of the human enteric nervous system  
30 (Pachnis, V. et al. (1998) Am. J. Physiol. 275:G183-G186).

#### 5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G.

et al. (1996) *J. Biol Chem.* 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised 5 of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

#### Kinases in Apoptosis

10 Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune disease, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain 15 appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) *J. Biol. Chem.*, 273:29066-29071). DRAK1 and DRAK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon- $\gamma$  induced 20 apoptosis (Sanjo et al. *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al. *supra*). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in 25 addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) *J. Biol. Chem.* 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation 30 and homeostasis of the immune system (Nagata, S. (1997) *Cell* 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase

catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 5 apoptosis pathway (Inohara et al. supra).

#### Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial 10 matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) *Adv. Enzyme Regul.* 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the 15 activity of the pyruvate dehydrogenase complex, which is an important regulatory enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) *Adv. Enzyme Regul.* 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate 20 dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) supra).

#### KINASES WITH NON-PROTEIN SUBSTRATES

25

#### Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) *Curr. Opin. Cell. Biol.* 30 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma

membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP<sub>2</sub>). PIP<sub>2</sub> is then cleaved into inositol triphosphate (IP<sub>3</sub>) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to 5 acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, 10 and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP<sub>2</sub>) to PI (3,4,5) P<sub>3</sub> (PIP<sub>3</sub>). PIP<sub>3</sub> then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, 15 and cdc42 (Shepherd, P.R., et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, supra).

PKC is also activated by diacylglycerol (DAG). Phorbol esters (PE) are analogs of DAG and 20 tumor promoters that cause a variety of physiological changes when administered to cells and tissues. PE and DAG bind to the N-terminal region of PKC. This region contains one or more copies of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. Diacylglycerol kinase (DGK), the enzyme that converts DAG into phosphatidate, contains two copies of the DAG/PE-binding domain in its N-terminal section (Azzi, A. et al. (1992) Eur. J. Biochem. 25 208:547-557).

An example of lipid kinase phosphorylation activity is the phosphorylation of D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has 30 emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including

platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al. *supra*).

#### Purine Nucleotide Kinases

5 The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of  
10 various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) *Cancer Res.* 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and  
15 utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R.J. et al. (1995) *J. Biol. Chem.* 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity to treat  
20 certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell  
25 receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21<sup>ras</sup> known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) *Cancer Res.* 49:4682-4689). High ratios of  
30 GTP:GDP caused by suppression of GuK cause activation of p21<sup>ras</sup> and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs

useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and Miller R.L. (1980) *J. Biol. Chem.* 255:7204-7207; Stenberg, K. et al. (1986) *J. Biol. Chem.* 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the  
5 necessary dosages of the drugs.

#### Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:11941-11945). Phosphorylation of  
10 deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for de novo synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and Eriksson, S. (1995) *Pharmacol. Ther.* 67:155-186).

The discovery of new human kinases, and the polynucleotides encoding them, satisfies a need  
15 in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human kinases.

20

#### SUMMARY OF THE INVENTION

The invention features purified polypeptides, human kinases, referred to collectively as "PKIN" and individually as "PKIN-1," "PKIN-2," "PKIN-3," "PKIN-4," "PKIN-5," "PKIN-6," "PKIN-7," "PKIN-8," "PKIN-9," "PKIN-10," "PKIN-11," "PKIN-12," "PKIN-13," "PKIN-14," "PKIN-15," "PKIN-16," "PKIN-17," "PKIN-18," "PKIN-19," "PKIN-20," "PKIN-21," and "PKIN-22." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of  
25 SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-22.  
30

The invention further provides an isolated polynucleotide encoding a polypeptide selected from

the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-22. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:23-44.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an

amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

The invention further provides an isolated polynucleotide selected from the group consisting of 5 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide 10 comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 15 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to 20 said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said 25 target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the 30 polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the

invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

5 The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide 10 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

15 The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide 20 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the 25 polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a 30 polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said

method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, ii) a 5 polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific 10 hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment 15 of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of 20 toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

25 Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for 30 analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and 5 polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these 10 may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a 15 reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. 20 Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the 25 invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"PKIN" refers to the amino acid sequences of substantially purified PKIN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

30 The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PKIN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

An "allelic variant" is an alternative form of the gene encoding PKIN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to 5 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PKIN include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PKIN or a 10 polypeptide with at least one functional characteristic of PKIN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PKIN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PKIN. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino 15 acid residues which produce a silent change and result in a functionally equivalent PKIN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PKIN is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and 20 arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, 25 polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. 30 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PKIN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small

molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. 5 Antibodies that bind PKIN polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used 10 carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies 15 which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX 20 (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include 25 deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

30 The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-

handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" 5 (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense 10 molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

15 The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PKIN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

20 "Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or 25 amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PKIN or fragments of PKIN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; 30 SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been

assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

5 "Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
10	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
15	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
20	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
25	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

30 Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

35 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains 40 at least one biological or immunological function of the natural molecule. A derivative polypeptide is

one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

5 "Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

10 "Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

15 A "fragment" is a unique portion of PKIN or the polynucleotide encoding PKIN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For 20 example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

25 A fragment of SEQ ID NO:23-44 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:23-44, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:23-44 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:23-44 from related polynucleotide sequences. The precise length of a fragment of SEQ ID 30 NO:23-44 and the region of SEQ ID NO:23-44 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-22 is encoded by a fragment of SEQ ID NO:23-44. A fragment of SEQ ID NO:1-22 comprises a region of unique amino acid sequence that specifically

identifies SEQ ID NO:1-22. For example, a fragment of SEQ ID NO:1-22 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-22. The precise length of a fragment of SEQ ID NO:1-22 and the region of SEQ ID NO:1-22 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the 5 intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two 10 or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore 15 achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in 20 Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

25 Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis 30 programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The

“BLAST 2 Sequences” tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

5        *Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

10      *Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, 15 over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

20        Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

25        The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

30        Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap

penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise 5 comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

10 *Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, 15 for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be 20 used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term ‘humanized antibody’ refers to an antibody molecule in which the amino acid 25 sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. 30 Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive

conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 5 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of 10 the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention 15 include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as 20 formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

25 The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_{0,t}$  or  $R_0$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate 30 to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune

disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PKIN which is 5 capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PKIN which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

10 The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PKIN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PKIN.

15 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

20 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

25 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

30 "Post-translational modification" of an PKIN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PKIN.

"Probe" refers to nucleic acid sequences encoding PKIN, their complements, or fragments

thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target 5 polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also 10 be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for 15 example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that 20 purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 25 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer 30 selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may

also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned 5 nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are 10 not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the 15 artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a 20 vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated 25 regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

30 An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PKIN, nucleic acids encoding PKIN, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

5 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the 10 epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

15 The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

20 "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

25 "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed 30 cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals

and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at

least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## 5 THE INVENTION

The invention is based on the discovery of new human human kinases (PKIN), the polynucleotides encoding PKIN, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders.

10 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is  
15 denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte 20 polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

25 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS 30 program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are human kinases.

For example, SEQ ID NO:1 is 91% identical to human casein kinase I-alpha (GenBank ID g852055) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The 5 BLAST probability score is 2.9e-167, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ 10 ID NO:1 is a protein kinase.

For example, SEQ ID NO:10 is 91% identical to Mus musculus FYVE finger-containing phosphoinositide kinase (GenBank ID g4200446) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also contains a 15 phosphatidyl inositol 4-phosphate 5-kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from PRODOM analysis provides further corroborative evidence that SEQ ID NO:10 is a phosphoinositide kinase.

For example, SEQ ID NO:12 is 71% identical to human serine/threonine protein kinase 20 (GenBank ID g7160989) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.7e-148, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden 25 Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:12 is protein kinase.

For example, SEQ ID NO:13 is 86% identical to murine pantothenate kinase 1 beta (GenBank ID g6690020) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.6e-129, which indicates the probability of obtaining the observed 30 polypeptide sequence alignment by chance. Pantothenate kinase (PanK) is proposed to be the master regulator of CoA biosynthesis in mammalian cells, by controlling flux through the CoA biosynthetic pathway. Changes in the level of tissue PanK activity is reflected by the concurrent changes in the levels of CoA as seen in various metabolic states. Alterations in CoA levels and PanK activity are

seen during starvation/feeding, pathological states such as diabetes and by treatment with hypolipidemic drugs (Rock, C.O. et al., (2000) *J. Biol. Chem.* 275:1377-1383.)

For example, SEQ ID NO:16 is 68% identical to *Mus musculus* Nck-interacting kinase-like embryo specific kinase (GenBank ID g6472874) as determined by the Basic Local Alignment Search 5 Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further 10 corroborative evidence that SEQ ID NO:16 is a protein kinase.

For example, SEQ ID NO:19 is 99% identical to human RET tyrosine kinase receptor (GenBank ID g5419753) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:19 also contains a eukaryotic 15 protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:19 is a tyrosine kinase.

For example, SEQ ID NO:22 is 33% identical to *Gallus gallus* smooth muscle myosin light 20 chain kinase precursor (GenBank ID g212661) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.2 e-60, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:22 also contains two eukaryotic protein kinase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family 25 domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:22 is a protein kinase.

SEQ ID NO:2-9, SEQ ID NO:11, SEQ ID NO:14-15, SEQ ID NO:17-18, and SEQ ID NO:20-21 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-22 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were 30 assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide

consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:23-44 or that distinguish between SEQ ID NO:23-44 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 183812R7 is the identification number of an Incyte cDNA sequence, and CARDNOT01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71583296V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation “ENST”). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation “NM” or “NT”) or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation “NP”). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a “stitched” sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, FLXXXXXX\_gAAAAAA\_gBBBBB\_1\_N is the identification number of a “stretched” sequence, with XXXXXX being the Incyte project identification number, gAAAAAA being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to

specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the “exon-stretching” algorithm, a RefSeq identifier (denoted by “NM,” “NP,” or “NT”) may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from 5 genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

15 In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library 20 is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PKIN variants. A preferred PKIN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid 25 sequence identity to the PKIN amino acid sequence, and which contains at least one functional or structural characteristic of PKIN.

The invention also encompasses polynucleotides which encode PKIN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:23-44, which encodes PKIN. The polynucleotide sequences 30 of SEQ ID NO:23-44, as presented in the Sequence Listing, embrace the equivalent RNA sequences,

wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PKIN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least 5 about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PKIN. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:23-44 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ 10 ID NO:23-44. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PKIN.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PKIN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, 15 the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PKIN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PKIN and its variants are generally capable of 20 hybridizing to the nucleotide sequence of the naturally occurring PKIN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PKIN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which 25 particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PKIN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PKIN and 30 PKIN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PKIN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:23-44 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PKIN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060).

Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National 5 Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include 10 sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary 15 sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer 20 controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PKIN may be cloned in recombinant DNA molecules that direct expression of PKIN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy 25 of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PKIN.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PKIN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA 30 shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or improve 5 the biological properties of PKIN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and 10 selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple 15 naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PKIN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, PKIN itself or a fragment thereof may be synthesized using chemical methods. For example, peptide 20 synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PKIN, or any part thereof, may be altered during direct synthesis and/or combined with sequences 25 from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. 30 (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active PKIN, the nucleotide sequences encoding PKIN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in

a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PKIN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PKIN. Such signals 5 include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PKIN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided 10 by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression 15 vectors containing sequences encoding PKIN and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

20 A variety of expression vector/host systems may be utilized to contain and express sequences encoding PKIN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or 25 tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New 30 York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola,

M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

5 In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PKIN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PKIN can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 10 plasmid (Life Technologies). Ligation of sequences encoding PKIN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of PKIN are needed, e.g. for the production of 15 antibodies, vectors which direct high level expression of PKIN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PKIN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such 20 vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of PKIN. Transcription of sequences 25 encoding PKIN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These 30 constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases

where an adenovirus is used as an expression vector, sequences encoding PKIN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PKIN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. 5 Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are 10 constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PKIN in cell lines is preferred. For example, sequences encoding PKIN can be transformed into cell 15 lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the 20 introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et 25 al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) 30 J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate

luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest 5 is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PKIN is inserted within a marker gene sequence, transformed cells containing sequences encoding PKIN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PKIN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates 10 expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PKIN and that express PKIN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or 15 chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PKIN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing 20 monoclonal antibodies reactive to two non-interfering epitopes on PKIN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) *Immunochemical Protocols*, Humana Press, 25 Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PKIN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. 30 Alternatively, the sequences encoding PKIN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety

of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

5        Host cells transformed with nucleotide sequences encoding PKIN may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PKIN may be designed to contain signal sequences which direct 10 secretion of PKIN through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, 15 lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture 20 Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

25        In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PKIN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PKIN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PKIN activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose 30 binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PKIN encoding sequence and the heterologous protein sequence, so that PKIN

may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PKIN may be achieved in 5 vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

PKIN of the present invention or fragments thereof may be used to screen for compounds 10 that specifically bind to PKIN. At least one and up to a plurality of test compounds may be screened for specific binding to PKIN. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PKIN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a 15 natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PKIN binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PKIN, either as a secreted 20 protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PKIN or cell membrane fractions which contain PKIN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PKIN or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is 25 detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PKIN, either in solution or affixed to a solid support, and detecting the binding of PKIN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural 30 product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PKIN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PKIN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PKIN

activity, wherein PKIN is combined with at least one test compound, and the activity of PKIN in the presence of a test compound is compared with the activity of PKIN in the absence of the test compound. A change in the activity of PKIN in the presence of the test compound is indicative of a compound that modulates the activity of PKIN. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PKIN under conditions suitable for PKIN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PKIN may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PKIN or their mammalian homologs may be 10 "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted 15 by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). 20 Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PKIN may also be manipulated in vitro in ES cells derived from 25 human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding PKIN can also be used to create "knockin" humanized animals 30 (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PKIN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with

potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PKIN, e.g., by secreting PKIN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

## THERAPEUTICS

5       Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PKIN and human kinases. In addition, the expression of PKIN is closely associated with brain, breast tumor, cardiovascular, digestive, fallopian tube tumor, fetal stomach, nervous, ovarian tumor, pancreatic tumor, peritoneal tumor, pituitary gland, placental, prostate tumor, neural, spinal cord, and testicular tissues, and with umbilical cord blood dendritic cells. Therefore, PKIN appears to play a  
10      role in cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders. In the treatment of disorders associated with increased PKIN expression or activity, it is desirable to decrease the expression or activity of PKIN. In the treatment of disorders associated with decreased PKIN expression or activity, it is desirable to increase the expression or activity of PKIN.

15       Therefore, in one embodiment, PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,  
20      gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune  
25      thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,  
30      myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral,

bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including

5 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy,

10 gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and

15 sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

20 degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung

25 anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic

30 pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and

noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

15 In another embodiment, a vector capable of expressing PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those described above.

20 In a further embodiment, a composition comprising a substantially purified PKIN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those provided above.

25 In still another embodiment, an agonist which modulates the activity of PKIN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those listed above.

30 In a further embodiment, an antagonist of PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN. Examples of such disorders include, but are not limited to, those cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders described above. In one aspect, an antibody which specifically binds PKIN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PKIN.

35 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the 5 various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PKIN may be produced using methods which are generally known in the art. In particular, purified PKIN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PKIN. Antibodies to PKIN may also 10 be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and 15 others may be immunized by injection with PKIN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG 20 (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PKIN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PKIN 25 amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PKIN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma 30 technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the

splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PKIN-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population 10 or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PKIN may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin 15 digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired 20 specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PKIN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PKIN epitopes is generally used, but a competitive binding assay may also be 25 employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PKIN. Affinity is expressed as an association constant, K<sub>a</sub>, which is defined as the molar concentration of PKIN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K<sub>a</sub> 30 determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PKIN epitopes, represents the average affinity, or avidity, of the antibodies for PKIN. The K<sub>a</sub> determined for a preparation of monoclonal antibodies, which are monospecific for a particular PKIN epitope, represents a true measure of affinity. High-affinity antibody preparations with K<sub>a</sub>

ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PKIN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_d$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PKIN, preferably in active form, from the antibody

5 (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a 10 polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PKIN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

15 In another embodiment of the invention, the polynucleotides encoding PKIN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PKIN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments 20 can be designed from various locations along the coding or control regions of sequences encoding PKIN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered 25 intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other 30 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PKIN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined 5 immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, 10 R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), 15 hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PKIN expression or regulation causes disease, the expression of PKIN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

20 In a further embodiment of the invention, diseases or disorders caused by deficiencies in PKIN are treated by constructing mammalian expression vectors encoding PKIN and introducing these vectors by mechanical means into PKIN-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and 25 (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

30 Expression vectors that may be effective for the expression of PKIN include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PKIN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter

(e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the 5 FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PKIN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver 10 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

15 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PKIN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PKIN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences 20 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. 25 (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. 30 Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998)

Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PKIN to cells which have one or more genetic abnormalities with respect to the expression of PKIN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PKIN to target cells which have one or more genetic abnormalities with respect to the expression of PKIN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PKIN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PKIN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based

on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PKIN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PKIN-coding RNAs and the synthesis of high levels of PKIN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of PKIN into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PKIN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for

secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared 5 by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PKIN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs 10 that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterate linkages 15 within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a 20 compound which is effective in altering expression of a polynucleotide encoding PKIN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective 25 compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PKIN expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PKIN may be therapeutically useful, and in the treatment of disorders associated with decreased PKIN expression or activity, a compound which specifically promotes expression of the 30 polynucleotide encoding PKIN may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in

altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a 5 polynucleotide encoding PKIN is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PKIN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the 10 polynucleotide encoding PKIN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide 15 can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified 20 oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. 25 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and 30 monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various

formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PKIN, antibodies to PKIN, and mimetics, agonists, antagonists, or inhibitors of PKIN.

The compositions utilized in this invention may be administered by any number of routes 5 including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. 10 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and 15 proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of 20 macromolecules comprising PKIN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PKIN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to 25 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for 30 administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PKIN or fragments thereof, antibodies of PKIN, and agonists, antagonists or inhibitors of PKIN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by

standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD<sub>50</sub>/ED<sub>50</sub> ratio. Compositions which exhibit large 5 therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

10 The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response 15 to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. 20 Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PKIN may be used for the 25 diagnosis of disorders characterized by expression of PKIN, or in assays to monitor patients being treated with PKIN or agonists, antagonists, or inhibitors of PKIN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PKIN include methods which utilize the antibody and a label to detect PKIN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and 30 may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PKIN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PKIN expression. Normal

or standard values for PKIN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PKIN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PKIN expressed in subject, 5 control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PKIN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect 10 and quantify gene expression in biopsied tissues in which expression of PKIN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PKIN, and to monitor regulation of PKIN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PKIN or closely related molecules may be used to 15 identify nucleic acid sequences which encode PKIN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PKIN, allelic variants, or related sequences.

20 Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PKIN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:23-44 or from genomic sequences including promoters, enhancers, and introns of the PKIN gene.

Means for producing specific hybridization probes for DNAs encoding PKIN include the 25 cloning of polynucleotide sequences encoding PKIN or PKIN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, 30 such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PKIN may be used for the diagnosis of disorders associated with expression of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,

teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder,

5 such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins,

10 erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,

15 thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and

20 cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker

25 muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly,

30 craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary

artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic

5 endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary

10 disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary

15 hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid

20 storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with

25 hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity. The polynucleotide sequences encoding PKIN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick,

30 pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PKIN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PKIN may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

sequences encoding PKIN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control 5 sample then the presence of altered levels of nucleotide sequences encoding PKIN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PKIN, 10 a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PKIN, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. 15 Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, 20 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or 25 overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PKIN 30 may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PKIN, or a fragment of a polynucleotide complementary to the polynucleotide encoding PKIN, and will be employed under optimized conditions for identification of a specific gene or

condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are 5 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal 10 tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP 15 (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the 20 high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PKIN include radiolabeling or biotinylation of nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be 25 accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray 30 can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor

progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and 5 display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PKIN, fragments of PKIN, or antibodies specific for PKIN may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

10 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 15 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The 20 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention 25 may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) 30 Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression

provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for

example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein 5 identification.

A proteomic profile may also be generated using antibodies specific for PKIN to quantify the levels of PKIN expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-10 111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and 15 should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid 20 degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. 25 A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological 30 sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the

two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PKIN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PKIN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery

techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant 5 invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PKIN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a 10 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PKIN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are 15 synthesized on a solid substrate. The test compounds are reacted with PKIN, or fragments thereof, and washed. Bound PKIN is then detected by methods well known in the art. Purified PKIN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

20 In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PKIN specifically compete with a test compound for binding PKIN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PKIN.

In additional embodiments, the nucleotide sequences which encode PKIN may be used in any 25 molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, 30 to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific

embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/242,410, U.S. Ser. No. 60/244,068, U.S. Ser. No. 60/245,708, U.S. Ser. 5 No. 60/247,672, U.S. Ser. No. 60/249,565, U.S. Ser. No. 60/252,730, and U.S. Ser. No. 60/250,807, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

10 Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl 15 cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles 20 (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the 25 UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300- 30 1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid

(Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or

5 ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an 10 AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a 15 high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## 20 III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the 25 MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the 30 ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the

techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The 5 Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.)

10 The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and 15 Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein 20 databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as 25 incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second 30 column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the

identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:23-44. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and 5 amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative human kinases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of 10 organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA 15 sequences encode human kinases, the encoded polypeptides were analyzed by querying against PFAM models for human kinases. Potential human kinases were also identified by homology to Incyte cDNA sequences that had been annotated as human kinases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to 20 the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted 25 coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### “Stitched” Sequences

30 Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm

based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to 5 be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants.

10 Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended 15 with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases 20 using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The 25 GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of PKIN Encoding Polynucleotides**

30 The sequences which were used to assemble SEQ ID NO:23-44 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:23-44 were assembled into clusters of contiguous and overlapping sequences using

assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment 5 of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in 10 humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease 15 genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:29 was mapped to chromosome 1 within the interval from 199.20 to 203.00 centiMorgans, to chromosome 13 within the interval from 105.20 centiMorgans to the q terminus, and to chromosome 6 within the interval from 59.60 to 72.20 centiMorgans. More than one map location is reported for SEQ ID NO:29, indicating that sequences having different map locations 20 were assembled into a single cluster. This situation occurs, for example, when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs 25 from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer 30 search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

BLAST Score x Percent Identity

$$5 \times \min\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PKIN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PKIN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### VIII. Extension of PKIN Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate

fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 5 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

10 High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg<sup>2+</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters  
15 for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

20 The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the 25 concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and 30 sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector

(Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

5 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA 10 recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

15 In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:23-44 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base 20 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a 25 SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: *Ase* I, *Bgl* II, *Eco* RI, *Pst* I, *Xba* I, or *Pvu* II (DuPont NEN).

30 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and

compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical 5 microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface 10 of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may 15 comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a 20 fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### 25 Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M 30 dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one

with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated 5 using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

#### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element 10 is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope 15 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

20 Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). 25 Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

30 Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly

larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

5 **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide 10 containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, 15 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

20 The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, 25 are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC 30 computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission

spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used 5 for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

## XI. Complementary Polynucleotides

Sequences complementary to the PKIN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PKIN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with 10 smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PKIN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PKIN-encoding transcript.

## 15 XII. Expression of PKIN

Expression and purification of PKIN is achieved using bacterial or virus-based expression systems. For expression of PKIN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid 20 promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PKIN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PKIN in eukaryotic cells is achieved by infecting insect or mammalian cell 25 lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PKIN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the 30 latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PKIN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PKIN at 5 specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified PKIN obtained by these methods can be used directly in the assays shown in 10 Examples XVI, XVII, and XVIII, where applicable.

### XIII. Functional Assays

PKIN function is assessed by expressing the sequences encoding PKIN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice 15 include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish 20 transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of 25 fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with 30 specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) *Flow Cytometry*, Oxford, New York NY.

The influence of PKIN on gene expression can be assessed using highly purified populations

of cells transfected with sequences encoding PKIN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success 5 NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PKIN and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XIV. Production of PKIN Specific Antibodies

PKIN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., 10 Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PKIN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for 15 selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to 20 increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PKIN activity by, for example, binding the peptide or PKIN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### 25 XV. Purification of Naturally Occurring PKIN Using Specific Antibodies

Naturally occurring or recombinant PKIN is substantially purified by immunoaffinity chromatography using antibodies specific for PKIN. An immunoaffinity column is constructed by covalently coupling anti-PKIN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is 30 blocked and washed according to the manufacturer's instructions.

Media containing PKIN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PKIN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt

antibody/PKIN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PKIN is collected.

## XVI. Identification of Molecules Which Interact with PKIN

PKIN, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. 5 (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PKIN, washed, and any wells with labeled PKIN complex are assayed. Data obtained using different concentrations of PKIN are used to calculate values for the number, affinity, and association of PKIN with the candidate molecules.

10 Alternatively, molecules interacting with PKIN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

15 PKIN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

## XVII. Demonstration of PKIN Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein 20 substrate by PKIN in the presence of gamma-labeled <sup>32</sup>P-ATP. PKIN is incubated with the protein substrate, <sup>32</sup>P-ATP, and an appropriate kinase buffer. The <sup>32</sup>P incorporated into the substrate is separated from free <sup>32</sup>P-ATP by electrophoresis and the incorporated <sup>32</sup>P is counted using a radioisotope counter. The amount of incorporated <sup>32</sup>P is proportional to the activity of PKIN. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid 25 analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma <sup>32</sup>P-ATP. Following the reaction, free avidin in solution is added for binding to the 30 biotinylated <sup>32</sup>P-peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma <sup>32</sup>P-ATP. The reservoir of the centrifuged unit containing the <sup>32</sup>P-peptide product as retentate is then counted in a scintillation counter. This procedure allows assay of any type of protein kinase

sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes are as follows: Histone H1 (Sigma) and p34<sup>cdc2</sup>kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, 5 Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods in Enzymology 200:62-81).

In another alternative, protein kinase activity of PKIN is demonstrated *in vitro* in an assay containing PKIN, 50 $\mu$ l of kinase buffer, 1 $\mu$ g substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10  $\mu$ g ATP, and 0.5 $\mu$ Ci [ $\gamma$ -<sup>33</sup>P]ATP. The reaction is 10 incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [ $\gamma$ -<sup>33</sup>P]ATP is removed by washing and the incorporated radioactivity is measured using a radioactivity scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS 15 loading buffer and visualized on a 12% SDS polyacrylamide gel by autoradiography. Incorporated radioactivity is corrected for reactions carried out in the absence of PKIN or in the presence of the inactive kinase, K38A.

In yet another alternative, adenylate kinase or guanylate kinase activity may be measured by the incorporation of <sup>32</sup>P from gamma-labeled <sup>32</sup>P-ATP into ADP or GDP using a gamma radioisotope counter. The enzyme, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and <sup>32</sup>P-labeled ATP as the phosphate donor. The 20 reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is cut out and counted. The radioactivity recovered is proportional to the enzyme activity.

In yet another alternative, other assays for PKIN include scintillation proximity assays (SPA), 25 scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of PKIN activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

Kinase activity of PKIN may be determined by its ability to convert polyphosphate substrate (PolyP) to ATP in the presence of ADP. PKIN and Poly P are incubated at 37°C for 40 minutes and 30 then at 90°C for 2 minutes in a buffer containing 50 mM Tris-HCl, pH 7.4, 40 mM ammonium sulfate, 4 mM MgCl<sub>2</sub>, and 5  $\mu$ M ADP. The reaction mixture is diluted 1:100 in 100 mM Tris-HCl (pH 8.0), 4 mM EDTA, which is then diluted 1:1 in luciferase reaction mixture (ATP Bioluminescence Assay Kit CLS II; Boehringer Mannheim). The ATP generated is then quantitated using a luminometer

(Kornberg, A. et al. (1999) *Annu. Rev. Biochem.* 68:89-125; Ault-Riché, D. et al. (1998) *J. Bacteriol.* 180:1841-1847).

Kinase activity of PKIN, as measured by phosphorylation of substrate, may be determined using an immune complex kinase assay well known in the art. COS7 cells are transfected with an 5 expression plasmid constructed from a FLAG tag expression vector (pME18S-FLAG) containing PKIN DNA. A control transfection using vector alone without the PKIN DNA insert is done in parallel. After 48 hours, the cells are lysed in buffer A (20 mM HEPES-NaOH, pH 7.5, 3 mM MgCl<sub>2</sub>, 100 mM NaCl, 1 mM dithiothreitol, 1 mM phenylmethanesulfonyl fluoride, 1 µg/ml leupeptin, 1 mM EGTA, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 10 mM NaF, 20 mM β-glycerophosphate, and 0.5% Triton X-100) and 10 centrifuged at 14,000 rpm. Supernatants are incubated with anti-FLAG antibody (M2 monoclonal antibody; Eastman Kodak Co.) in a 50% slurry of protein A-Sepharose (Amersham Pharmacia Biotech) for 1.5 hours at 4°C. Immune complexes are precipitated and washed twice in buffer A and twice in buffer B (20 mM HEPES-NaOH, pH 7.5, 1 mM dithiothreitol, 10 µM Na<sub>3</sub>VO<sub>4</sub>, 2 mM β-glycerophosphate, 0.1 mM phenylmethanesulfonyl fluoride, 0.1 µg/ml leupeptin, 0.1 mM EGTA.) 15 Precipitates are incubated in buffer B containing 0.17 mg/ml myelin basic protein (MBP) (Sigma), 20 µM ATP, and 5 µCi of [γ-<sup>32</sup>P]ATP (NEN Life Science Products) at 30°C for 20 minutes. The reaction is stopped by the addition of 4X Laemmli sample buffer (50 mM Tris-HCl, pH 6.8, 2% SDS, 30 mM dithiothreitol, and 10% glycerol) and heated at 95°C for 5 minutes. Proteins are separated by SDS-polyacrylamide gel electrophoresis and radioactivity incorporated into MBP is detected by 20 autoradiography (Nakano, K. et al. (2000) *J. Biol. Chem.* 275:20533-20539.)

In yet another alternative, an assay for PanK activity of PKIN includes the enzyme preparation method as described in Vallari, D.S. et al., (1987) *J. Biol. Chem.* 262:2468-247. Pantothenate kinase-specific activities in cell lysates are calculated as a function of protein concentration with the assay being linear with respect to both time and protein input. Protein 25 concentrations are measured using the Bradford assay using bovine γ-globulin as a standard. Standard assays contain D-[1-<sup>14</sup>C]pantothenate (45.5 µM; specific activity 55 mCi/mmol), ATP (2.5 mM, pH 7.0), MgCl<sub>2</sub> (2.5 mM), Tris-HCl (0.1 M, pH 7.5), and 15 µg of protein from a soluble cell extract in a total volume of 40 µl. The mixture is incubated for 10 min. at 37 °C, and the reaction is stopped by depositing a 30-µl aliquot onto a Whatman DE81 ion-exchange filter disc which is then 30 washed in three changes of 1% acetic acid in 95% ethanol (25 ml/disc) to remove unreacted pantothenate. 4'-Phosphopantothenate is quantitated by counting the dried disc in 3 ml of scintillation solution (Rock, supra).

## XVIII. Enhancement/Inhibition of Protein Kinase Activity

Agonists or antagonists of PKIN activation or inhibition may be tested using assays described in section XVII. Agonists cause an increase in PKIN activity and antagonists cause a decrease in PKIN activity.

5        Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious  
10      to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incite Project ID	Polypeptide SEQ ID NO:	Incite Polypeptide ID	Polynucleotide SEQ ID NO:	Polynucleotide ID
7482896	1	7482896CD1	23	7482896CB1
7483046	2	7483046CD1	24	7483046CB1
71636374	3	71636374CD1	25	71636374CB1
7480597	4	7480597CD1	26	7480597CB1
3227248	5	3227248CD1	27	3227248CB1
4207273	6	4207273CD1	28	4207273CB1
7483334	7	7483334CD1	29	7483334CB1
7483337	8	7483337CD1	30	7483337CB1
6035509	9	6035509CD1	31	6035509CB1
7373485	10	7373485CD1	32	7373485CB1
5734965	11	5734965CD1	33	5734965CB1
7473788	12	7473788CD1	34	7473788CB1
3107989	13	3107989CD1	35	3107989CB1
7482887	14	7482887CD1	36	7482887CB1
2963414	15	2963414CD1	37	2963414CB1
7477139	16	7477139CD1	38	7477139CB1
55009053	17	55009053CD1	39	55009053CB1
7474648	18	7474648CD1	40	7474648CB1
7483053	19	7483053CD1	41	7483053CB1
7483117	20	7483117CD1	42	7483117CB1
7484498	21	7484498CD1	43	7484498CB1
7638121	22	7638121CD1	44	7638121CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Incyte Polypeptide NO:	GenBank NO:	ID	Probability Score	GenBank Homolog
1	7482896CD1	g852055		2.90E-167		[Homo sapiens] casein kinase I-alpha Fish,K.J. et al., (1995) J. Biol. Chem. 270:14875-14883
2	7483046CD1	g2736151		4.20E-167		[Rattus norvegicus] myotonic dystrophy kinase-related Leung,T. et al., (1998) Mol. Cell. Biol. 18:130-140
3	71636374CD1	g7549223	0			[Mus musculus] PALS1 (proteins associated with Lin-7, a membrane-associated guanylate kinase) Kamberov,E. et al., (2000) J. Biol. Chem. 275:11425-11431
4	7480597CD1	g2224679		1.40E-97		[Homo sapiens] KIAA0369 doublecortin-like kinase Nagase,T. et al., (1997) DNA Res. 4:141-150 Burgess,H.A. et al. (1999) J. Neurosci. Res. 58:567-575
5	3227248CD1	g6690020		4.90E-199		[Mus musculus] pantothenate kinase 1 beta Rock,C.O. et al. (2000) J. Biol. Chem. 275:1377-1383
6	4207273CD1	g4028547		4.70E-68		[Dictyostelium discoideum] MEK kinase alpha Chung,C.Y. et al. (1998) Genes Dev. 12:3564-3578
7	7483334CD1	g479173		1.70E-251		[Homo sapiens] protein kinase Schultz,S.J. et al. (1994) Cell Growth Differ. 5:625-635
8	7483337CD1	g9280288		3.10E-27		[Arabidopsis thaliana] receptor protein kinase Kaneko,T. et al. (2000) DNA Res. 7:217-221
9	6035509CD1	g6110362		3.60E-76		[Homo sapiens] Traf2 and NCK interacting kinase, splice variant 7 Fu,C.A. et al. (1999) J. Biol. Chem. 274:30729-30737
10	7373485CD1	g4200446	0			[Mus musculus] FYVE finger-containing phosphoinositide kinase Shisheva,A. et al. (1999) Mol. Cell. Biol. 19:623-634
11	5734965CD1	g2905643		4.60E-109		[Klebsiella pneumoniae] ribitol kinase Heuel,H. et al. (1998) Microbiology 144(Pt 6):1631-9
12	7473788CD1	g7160989		1.70E-148		[Homo sapiens] serine/threonine protein kinase Ruiz-Perez VL, et al. (2000) Nat. Genet. 24(3):283-6
13	3107989CD1	g6690020		1.60E-129		[Mus musculus] pantothenate kinase 1 beta Rock,C.O. et al. (2000) J. Biol. Chem. 275:1377-1383

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank NO:	ID	Probability Score	GenBank Homolog
14	7482887CD1	g205662	3.90E-48		[Rattus norvegicus] nucleoside diphosphate kinase Kimura, N. et al. J. Biol. Chem. (1990) 265:15744-15749
15	2963414CD1	g6524024	8.90E-106		[Mus musculus] mammalian inositol hexakisphosphate kinase 1 Saiardi, A. et al. Curr. Biol. (1999) 9:1323-1326
16	7477139CD1	g6472874	0		[Mus musculus] Nck-interacting kinase-like embryo specific kinase Nakano, K. et al. J. Biol. Chem. (2000) 275:20533-20539
17	55009053CD1	g15131540	0		[fl][Homo sapiens] (AJ316534) serine/threonine protein kinase
18	7474648CD1	g14346040	0		[fl][Homo sapiens] serine/threonine kinase PSKH2
19	7483053CD1	g5419753	0		[Homo sapiens] RET tyrosine kinase receptor Bordeaux, M.C. et al. (2000) EMBO J. 19:4056-4063
20	7483117CD1	g644770	2.70E-136		[Xenopus laevis] Wee1A kinase Mueller, P.R. et al. (1995) Mol. Biol. Cell 6:119-134
21	7484498CD1	g3599509	0		[Mus musculus] rho/rac-interacting citron kinase Di Cunto, F. et al. (1998) J. Biol. Chem. 273:29706-29711
22	7638121CD1	g212661	1.20E-60		[Gallus gallus] smooth muscle myosin light chain kinase precursor Olson, N.J. et al. (1990) Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7482996CD1	337	S105 S122 S199 S237 S242 S27 S49 S7 S96 T109 T146 T184 T228 T243 T323 T327 T38 Y209 Y274	N167 N215 N3	Eukaryotic protein kinase domain: Y17-F211 Protein kinases signatures and profile: T112-R168	HMMER-PFAM PROFILE-SCAN
2	7483046CD1	475		S161 S280 S307 S363 S407 S430 T455	Protein KINASE DOMAIN DM00004 P35506 19-273: L19-Y274 P54367 22-276: L19-Y274 P48730 11-265: L19-Y274 B56406 19-273: L19-Y274 CASEIN KINASE I TRANSFERASE SERINE/THREONINE PROTEIN ATP-BINDING ISOFORM ALPHA CK1 ALPHA MULTIGENE PD006522: R282-G324 Tyrosine kinase catalytic domain PR00109: Y126-M144 Kinase Protein Domain PD00534: V20-G29 Protein kinases ATP-binding region signature: I23-K46 Serine/Threonine protein kinases active-site signature: F132-M144 signal_cleavage: M1-G40 Eukaryotic protein kinase domain: F71-F337	BLAST- DOMO BLAST- PRODOM BLIMPS- PRINTS BLIMPS- PRODOM MOTIFS MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN KINASE DOMAIN DM000041 Q09013 83-336: I73-R325 S42867 75-498: I73-H252 I38133 90-369: E72-L220 P53894 353-658: L74-G215  KINASE PHORBOLESTER BINDING DYSTROPHIN KINASE RELATED CDC42 BINDING SIMILAR SERINE/THREONINE PROTEIN GENGHIS KHAN PD012280: L25-D70 Tyrosine kinase catalytic domain PR00109: M148- S161, S185-L203, C257-E279 Protein kinase C terminal domain: P351-D366  Protein kinases ATP-binding region signature: I77- K100 Serine/Threonine protein kinases active-site signature: Y191-L203  signal_cleavage: M1-S37  Guanylate kinase: T515-I624	BLAST- DOMO  BLAST- PRODOM  BLIMPS- PRINTS HMMER- PFAM  MOTIFS  SPSCAN HMMER- PFAM
3	71636374CD1	675	S130 S14 S143 S25 S383 S432 S517 S562 S569 S576 S581 S646 S84 T137 T253 T270 T422 T465 T514 T558 T584 T97 Y593	N82	GUANYLATE KINASE DM00755 A57653 370-570: P475-P670 P54936 769-955: R478-P670 I38757 709-898: Q474-P670 P310 6 529-718: R480-P670	BLAST- DOMO

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN DOMAIN SH3 KINASE GUANYLATE TRANSFERASE ATPBINDING REPEAT GMP MEMBRANE PD001338: T514-E620	BLAST- PRODOM
					SIMILAR TO GUANYLATE KINASE PD065809: G41-Q337	BLAST- PRODOM
					Guanylate kinase protein BL00856: V511-Y531, D539-R586	BLIMPS- BLOCKS
					SH3 domain signature PR00452: D386-E395, I348-P358, L369-Q384	BLIMPS- PRINTS
					PDZ domain (Also known as DHR or GLGF). PDZ: I256-S335,	HMMER- PFAM
					SH3 domain SH3:I348-Q415	HMMER- PFAM
					ATP/GTP-binding site motif A (P-loop): A404-S411	MOTIFS
					Guanylate Kinase signature and profile: T514-V531	MOTIFS
4	7480597CD1	835	S111 S153 S174 S223 S249 S271 S292 S349 S369 S380 S389 S393 S405 S525 S54 S59 S633 S713 T129 T194 T246 T278 T300 T319 T33 T451 T477 T499 T514 T545 T610 T63 T681 T790 T808	N768	Eukaryotic protein kinase domain pkinase: Y543-I800 Protein kinases signatures and profile: D640-I697	HMMER- PFAM PROFILE- SCAN
					PROTEIN KINASE DOMAIN DM00004 S57347 21-266: V548-T790 P08414 44-285: I549-T790 A44412 16-262: I549-A791 JU0270 16-262: I549-A791	BLAST- DOMO

Table 3

SEQ ID	Incyte Polypeptide ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: E609-V693	BLAST-PRODOM
					Octicosapeptide repeat pPF00564: Y543-S597, H605-M655, K473-G526	BLIMPS-PFAM
5	3227248CD1	373	S100 S283 S285 S330 S47 T10 T167 T209 T226 T230 T244 T34	N103 N72	Tyrosine kinase catalytic domain PR00109: L618-1631, H654-V672	BLIMPS-PRINTS
6	4207273CD1	735	S100 S111 S113 S124 S152 S170 S179 S185 S186 S20 S202 S215 S221 S225 S240 S267 S271 S302 S459 S503 S729 S9 T10 T105 T13 T30 T402 T417 T425 T469 T626 T663 T669 T84 Y512	N289 N312 N341 N392 N400 N61 N624 N647	PROTEIN KINASE DOMAIN DM000004 A48084 98-348:K470-A722 DM000004 Q01389 1176-1430:K470-A722 DM000004 P41892 11-249:G471-R719 DM000004 Q10407 826-1084:K470-A722	BLAST-DOMO
					KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001:L631-P673, E472-C537, Y533-S633, S701-S734	BLAST-PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7483334CD1	506	S148 S206 S243 S319 S325 S354 S47 T197 T288 T293 T308 T321 T373 T386 T402 T403 T479	N181 N345 N377 N401	Tyrosine kinase catalytic domain signature PR00109:M547-N560, Y583-L601, G636-I646, S655- M677 Eukaryotic protein kinase domain kinase:W468- L731 Protein_Kinase_Atp L474-K496 Protein_Kinase_St V589-L601 Protein kinases signatures and profile protein_kinase_tyr.prf:V569-A619	BLIMPS- PRINTS HMMER- PFAM MOTIFS MOTIFS PROFILE- SCAN BLAST- DOMO
8	7483337CD1	2014	S1076 S1151 S1177 S1217 S1274 S1279 S1454 S15 S1515 S1679 S1700 S1811 S1833 S1887 S1890 S1999	N1024 N1119 N1338 N1599 N1674 N307 N371 N409	Tyrosine kinase catalytic domain signature PR00109:M79-K92, H117-L135, S183-N205, Y226- A248 Eukaryotic protein kinase domain kinase:Y4-V257 Protein_Kinase_Atp I110-K33 Protein_Kinase_St V123-L135 Protein kinases signatures and profile protein_kinase_tyr.prf:M103-M156	BLIMPS- PRINTS HMMER- PFAM MOTIFS MOTIFS PROFILE- SCAN BLAST- DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			S203 S25 S321 S337 S401 S531 S56 S565 S599 S81 S843 S863 S887 S900 T1091 T1099 T1113 T1187 T1189 T1234 T1401 T1543 T1605 T1634 T1660 T1872 T1895 T2010 T280 T494 T517 T524 T533 T537 T680 T687 T699 T702 T703 T753 T795 T811 T835 T909 Y1225 Y1997 Y907		Tyrosine kinase catalytic domain signature PR00109: Y1414-V1432, V1483-H1505, Q1529-A1551 transmembrane domain transmem_domain:P1367-N1387 Eukaryotic protein kinase domain pkinase:E1280-P1549 Protein kinases signatures and profile protein_kinase_tyr.prf:L1400-E1457 Atp_Gtp_A_G672-S679	BLIMPS-PRINTS HMMER HMMER-PFAM PROFILE-SCAN MOTIFS
9	6035509CD1	348	S101 S171 S199 S271 S50 S7 T178 T213 T311 T318 T33	N177	PROTEIN KINASE DOMAIN DM00004 P10676 18-272:I17-P270 A5371 4 17-262:I17-S271 P38692 24-266:E19-S271 P08458 20-262:I21-S271	BLAST-DOMO
10	7373485CD1	2042	S1020 S105 S1079 S1125 S1130 S1148 S13	N1061 N1274 N1647 N1671	Tyrosine kinase catalytic domain signature PR00109: H134-L152, G181-I191, W250-V272 Eukaryotic protein kinase domain signature:W15-1281 Protein_Kinase_Atp I21-K44 Protein_Kinase_St I140-L152 Protein kinases signatures and profile protein_kinase_tyr.prf:M120-T172 Probable phosphatidyl inositol 4-phosphate 5-kinase FAB1 EC 2.7.1.68 1-phosphatidyl inositol 4-	BLIMPS-PRINTS HMMER-PFAM MOTIFS MOTIFS PROFILE-SCAN BLAST-PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			S1377 S1419 S1429 S1440 S1466 S1483 S1488 S1544 S1545 S1620 S1639 S1648 S168 S1685 S1703 S1784 S1785 S1830 S1899 S228 S244 S257 S261 S286 S291 S367 S423 S475 S502 S576 S789 S810 S835 S85 S872 S896 S977 T1005 T1013 T109 T1149 T1295 T1386 T1524 T1567 T1670 T1674 T1681 T1708 T1722 T173 T1743 T1813 T1852 T1872 T1909 T1970 T341 T342 T591 T666 T731 T782 T976 T984 Y1290 Y1716Y1933 Y659	N1870 N303 N310 N333	phosphate 5-kinase diphosphoinositide transferase PD136025:H461-F821, W1147-K1437, L1375- S1702, K638-K767, P1663-V1780, D1372-Q1458, F959-I1069, R960-D1053, F200-R262, D1895- S1950; PD041996-L1974-W2035	BLAST- PRODOM
					5-KINASE PHOSPHATIDYL INOSITOL 4- PHOSPHATE KINASE TYPE TRANSFERASE DIPHOSPHOINOSITIDE 1-PHOSPHATIDYL INOSITOL 4-PHOSPHATE II ALPHA PHOSPHATIDYL INOSITOL PD002308:PI751- G1966, L1974-F2028, 1493-H533	
					FYVE zinc finger FYV:Q153-C213 Phosphatidylinositol-4-phosphate 5-Kinase PIP5:R1791-F2028	MOTIFS
						MOTIFS
11	5734965CD1	551	S107 S176 S2 S21 S257 S368 S502 S54 T183 T286 T334 T356 T403 T66 Y526 Y531	N127 N219	FGGY family of carbohydrate kinases: L423-A490 FGGY FAMILY OF CARBOHYDRATE KINASES DM01757 P219391-480: V13-A184 XYLULOKINASE DM02388 P181571-492: T383- E539	HMMER- PFAM BLAST- DOMO BLAST- DOMO

Table 3

SEQ	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					FGGY FAMILY OF CARBOHYDRATE KINASES DM01757 P37677 I-479: R10-D260 FGGY FAMILY OF CARBOHYDRATE KINASES DM01757 P46834 I-488: Y11-V268 MPA43 PROTEIN PD130314:V13-I210	BLAST- DOMO BLAST- DOMO BLAST- PRODOM BLIMPS- BLOCKS
12	7473788CD1	485	S10 S159 S3 S343 S362	N405	Eukaryotic protein kinase domain: F93-Q345	HMMER- PFAM
13	3107989CD1	282	S415 S417 T115 T192 T466 T469 T76 Y119		PROTEIN KINASE DOMAIN DM00004 P54644 I122- 362: I95-S342 PROTEIN KINASE DOMAIN DM00004 P28178 I155- 393: I95-L341 PROTEIN KINASE DOMAIN DM08046 P05986 I-397: K65-P372 P06244 I-396: F93-P372 Tyrosine kinase catalytic domain signature PR0109:V170-Q183, Y206-L224 Protein kinases ATP-binding region signature I99- K122 Serine/Threonine protein kinases active-site signature I212-L224 signal_cleavage:M1-A24 signal_cleavage: M1-A27	BLAST- DOMO BLAST- DOMO BLAST- DOMO BLIMPS- PRINTS MOTIFS MOTIFS SPSCAN SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7482887CD1	151	S42 S97 T35 Y141		NUCLEOSIDE DIPHOSPHATE KINASES DM00773P488173-152:17-Y150 DM00773I39074119-168:17-Y150 DM00773Q076611-148:17-Y150 DM00773P505901-150:17-Y150 KINASE DIPHOSPHATE NUCLEOSIDE TRANSFERASE NDK NDP ATP-BINDING PROTEIN I PRECURSOR PD001018:17-Y150 Nucleoside diphosphate kinases proteins BL00469:E77-L131	BLAST- DOMO
15	2963414CD1	410	S134 S156 S276 S318 T259 T361 T374 T383 T62	N117 N290	ndp_kinases:G96-R142 PROTEIN ARGININE METABOLISM REGULATION III TRANSCRIPTION SIMILARITY SACCHAROMYCES CEREVISIAE PUTATIVE PD011544:S188-Q333, S355-L403 PUTATIVE BZIP TRANSCRIPTION FACTOR CHROMOSOME IV READING FRAME ORF YDR017C PD024140:G15-R197 Aldo/keto reductase family putative active site signature I312-L327	BLAST- PRODOM BLAST- PRODOM BLIMPS- BLOCKS HMMER- PFAM PROFILE- SCAN BLAST- PRODOM

Table 3

SEQ ID NO:	Incute Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	7477139CD1	1581	S101 S1107 S1112 S1139 S1178 S1233 S1291 S1346 S136 S1400 S1426 S1435 S148 S1537 S1577 S211 S283 S376 S498 S580 S671 S676 S700 S709 S718 S749 S807 S84 S890 S891 S892 S910 T1071 T1123 T1194 T1367 T1508 T1546 T1556 T246 T276 T294 T357 T573 T664 T690 T899 T981 T992	N1137 N1201 N146 N654 N668 N990	PROTEIN KINASE DOMAIN DM00004 P10676 18-272:Y83-P302 DM00004 A53714 17-262:L43-S304 DM00004 P38692 24-266:S84-C293, K29-N57 DM00004 P50327 388-627:K77-S304, I31-E65	BLAST_DO MO
17	55009053CD1	1084	S1024 S1031 S1038 S1042 S1058 S157 S172 S231 S25 S422 S452 S478 S52 S521 S552 S569 S604 S623 S709 S80 S862 S882 S895 S914 S962 S968 S969 S981 S988 T102 T1037 T167 T230 T256 T263 T37 T420 T48 T543 T593 T631 T8 Y1005	N953	Serine/Threonine protein kinases active-site signature 1139-1151 Protein kinases signatures and profile protein_kinase_tyr.prf: L118-F173 Eukaryotic protein kinase domain pkinase: L15-F273	MOTIFS
18	7474648CD1	600	S206 S331 S369 S425 S456 S543 S55 S571	N18 N495	Protein kinases ATP-binding region signature I284- K307	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			S577 S585 T117 T14 T25 T299 T300 T356 T371 T395 T433 T58		Eukaryotic protein kinase domain pkinase: Y278-V535 Tyrosine kinase catalytic PR00109: M352-I365, Y388-BLIMPS- Y406, V458-E480	HMMER- PFAM
19	7483053CD1	1114	S1034 S104 S110 S131 S159 S173 S224 S363 S413 S457 S522 S561 S65 S670 S691 S696 S765 S811 S819 S836 S922 T1022 T1055 T1078 T261 T295 T315 T328 T350 T456 T492 T538 T564 T675 T729 T75 T847 T930 Y1096 Y483 Y905	N1092 N151 N199 N336 N343 N361 N367 N377 N394 N448 N468 N554 N834 N975 N98 K758	PROTEIN KINASE DOMAIN DM00004 S57347 21- 266: D279-L516 P08414 44-285: I280-S525 JN0323 25-268: I284-R523 S46284 28-274: I284- A526  signal peptide: M1-G28 Signal cleavage: M1-A26 Transmembrane domain: L13-F31 Protein kinases ATP-binding region signature L730- K758 Tyrosine protein kinases specific active-site signature L870-V882 Protein kinases signatures and profile protein kinase, tyr:prf: D850-D903 Receptor tyrosine kinase class II signature receptor_tyr_kin_ii:prf: R878-D925 Cadherin domain cadherin: P172-T261  Eukaryotic protein kinase domain pkinase: L724- L1005	BLAST- DOMO  HMMER- SPSCAN HMMER MOTIFS MOTIFS MOTIFS HMMER MOTIFS PROFILE- SCAN PROFILE- SCAN HMMER- PFAM HMMER- PFAM

Table 3

SEQ	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Receptor tyrosine kinase BL00239: D903-Y952, P957-I1001, E775-V822, L851-R873, A876-E901 BL00240: K716-A764, A764-E818, D850-K887, E902-G949, G949-I1001 BL00790: G748-L801, A855-A876, A877-D903, Q910-W942, H968-L1016	BLIMPS-BLOCKS
					Tyrosine kinase catalytic PR00109: V804-R817, Y864-V882, I913-L923, S932-G954, C976-F998	BLIMPS-PRINTS
20	7483117CD1	567	S162 S17 S206 S243 S278 S543 S552 S70 T112 T125 T22 T246 T544 T559 T68 Y238	N15 N332	RECEPTOR KINASE PRECURSOR SIGNAL RET TYROSINE PROTOONCOGENE TYROSINE CRET TRANSFERASE PD014372; P273-K666, D300-V725; PD014143; Y30-C197; PD007958: V1010-G1063, PD010335; M1064-S1114	BLAST-PRODOM
					PROTEIN-TYROSINE KINASE RET DM05080 P07949 302-723: D302-L724 I48735 303-724: D302-L724 PROTEIN KINASE DOMAIN DM00004 JN0290 88-360: V725-F998 P07949 725-997: V725-F998	BLAST-DOMO
					Protein kinases ATP-binding region signature I218-K241	MOTIFS
					Serine/Threonine protein kinases active-site signature M335-I347	MOTIFS
					Eukaryotic protein kinase domain kinase F212-L480	HMMER-PFAM
					Tyrosine kinase catalytic site PR00109: N289-S302, Y329-I347, A415-G437, L455-A477	BLIMPS-PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases		
21	748498CD1	2054	S81 S93 S140 S248 S308 S361 S381 S386 S410 S436 S445 S480 S487 S501 S516 S529 S546 S577 S582 S699 S883 S888 S924 S1031 S1049 S1097 S1158 S1160 S1234 S1315 S1364 S1365 S1370 S1371 S1377 S1574 S1845 S1915 S1933 S2014 S2028 T83 T378 T498 T604 T840 T951 T956 T989 T1041 T1062 T1112 T1186 T1231 T1309 T1326 T1336 T1372 T1543 T1583	N835 N1622 N1745 N1768	WEE1 HOMOLOG WEE1-LIKE PROTEIN KINASE MITOSIS TRANSFERASE TYROSINEPROTEIN ATPBINDING PHOSPHORYLATION PD028078: N483_G561	PROTEIN KINASE DOMAIN DM000004 P47817 211-470: L213-A477 P30291 300-559: E214- A477 P54350 241-507: E214-A477 A57247 104- 343: K217-I347, A366-R474	CNH (NIK-1 like kinase) domain: L1619-Y1916	BLAST- PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			T1775 T1787 T1943 T1955 T1961 T2015 Y763		CITRON PROTEIN COILED COIL RHO/RACINTERACTING KINASE PD155701: F859-L1071 PD143273: G1439-V1631 PD082663: L1201-P1389 PD143272: A1881-V2054	BLAST- PRODOM
					PROTEIN KINASE DOMAIN DM00004 Q09013 83-336: V99-L349 S42867 75-498: S101-G241, I258-S445 S42864 41-325: E98-G241, N249-L349 P53894 353-658: L102-G241 I258-L349	BLAST- DOMO
22	7638121CD1	1665	S97 S152 S156 S163 S242 S364 S450 S459 S491 S493 S528 S536 S588 S762 S827 S875 S915 S917 S929 S947 S961 S997 S1087 S1147	N1005	Protein kinases ATP-binding region signature V103-K126 Serine/Threonine protein kinases active-site signature: Y217-V229 Leucine zipper pattern: L854-L875, L991-L1012, L1057-L1078, L1159-L1180	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			S1203 S1336 S1351 S1365 S1391 S1434 S1446 S1459 S1461 S1521 T59 T230 T257 T312 T668 T870 T965 T1211 T1310 T1320 T1638		Tyrosine kinase catalytic domain signature PR00109: S341-E363, L387-A409, L238-Y251, Y274-M292	BLIMPS-PRINTS
					KINASE PROTEIN TRANSFERASE ATPBINDING SERINE/THREONINEPROTEIN PHOSPHORYLATION RECEPTOR TYROSINEPROTEIN PRECURSOR TRANSMEMBRANE PD000001: V256-V327, S323-D365, S380-P423	BLAST-PRODOM
					PROTEIN KINASE DOMAIN DM00004 JN0583 727-969: V167-R401, Q1372-P1563 P07313 298-541: K168-A409, Q1378-P1563 P53355 15-257: E169-R406, Q1374-P1563 S07571 5152-5396: E166-R406, Q1374-P1606	BLAST- DOMO
					Protein kinases ATP-binding region signature I171-K194	MOTIFS
					Tyrosine protein kinases specific active-site signature I1484-I1496	MOTIFS
					Protein kinase St V280-M192	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
23	7482896CB1	1014	982-1014	GNN.g7899226_000043_002 .edit	1	1014
24	7483046CB1	1530	719-770, 1-61, 1036-1104, 1271-1461, 313-464	71583296V1 71581650V1 71601507V1 55143579J1 71579961V1 55140831J1	889 778 1124 1 266 118	1476 1455 1530 272 884 522
25	71636374CB1	3150	1294-1806, 1-115, 2593-2616	183812R7 (CARDNOT01) 7676860H1 (NOSETUE01) 8252304H1 (BRAHDIT10) 5223511F9 (OVARDIT07)	2581 250 25 1225	3148 864 804 1397
				GBI.g7452884_edit GBI.g8919852_edit	1125 1099	2085 1898
				7214961H1 (LUNGFEC01) 7710619J1 (TESTTUE02) 7391509H1 (LIVRFE02) 5958404H1 (BRATNOT05)	1 1611 751 2796	250 2273 1302 3150
				5971916H1 (BRAZNOT01)	2211	2832
26	7480597CB1	2901	1907-1981, 1-156, 748-1606, 255-313	55150024J1 5507363J1J1 55150108J1 2841337T6 (DRGLNOT01) 55144761T1 5543295F7 (TESTNOC01) GNN.g7658410_000016_002 1 56001404J1	1377 630 1711 2251 2132 137 1 1790	2056 1518 2070 2901 2833 574 2013 2434

Table 4

Polynucleotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
27	3227248CB1	1671	1-85,	70944845V1	997	1646
			1593-1671,	7207691H1 (F1BPFEA01)	451	1050
			1327-1360	82283762T1 (L1VRRNON08)	180	562
				GB1_g9796547_edit	1	1539
				71281138V1	1089	1671
				5260904F6 (CONDITUT01)	569	1065
28	4207273CB1	2577	1-1641,	5543515F6 (TESTNOC01)	907	1376
			1845-1889	5357164H1 (TESTNOC01)	238	440
				55144823H1	2112	2577
				GNN_g9230839_0000001_002	1	1293
				55073166J1	1115	1773
				4919888T6 (TESTNOT11)	1445	2141
29	7483334CB1	2110	1-640,	71341632V1	1559	2110
			1255-1314,	71341335V1	1145	1708
			948-1005	940589R6 (ADRENOT03)	1916	2110
				6512850H1 (THYMDIT01)	1007	1688
				6102073H1 (UTRENO109)	797	1087
				4970029F7 (KIDEUNC10)	1	677
				7659406H1 (OVARNOE02)	509	1081
30	7483337CB1	7093	1-3002,	7383958R8 (FTUBTUE01)	1	694
			4789-5840,	3245534H1 (BRAINNOT19)	2681	2928
			7069-7093,	72334852V1	5219	5761
			3561-3671	7383958F8 (FTUBTUE01)	537	1196
				58002303T1	6221	7093
				70771904V1	5851	6475
				GNN_g6693375_0000016_002	986	3303
				.edit		
				55046508H1	2906	3666

Table 4

Polynucleotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
			55144427J1		5514	6397
			5208289H1 (BRAFNOT02)	4900	5138	
			7036825F6 (UTRSTM02)	3953	4647	
			55046508J1	3448	4132	
			70772942V1	5079	5680	
			6436908H1 (LUNGNON07)	908	1407	
			GNN g6721428_000012_004	3780	6267	
			.edit			
31	603509CB1	1800	152-333, 1-25,	71927475V1	1340	1800
			1463-1800, 770-862	6035509F8 (PITUNOT06) 55071284J1 72420180D1 55071288J1	848 818 1 480	1614 818 1098 1096
					2075	2717
32	7373485CB1	6347	4445-5413, 728-786, 6321-6347, 1497-3441, 4019-4079, 877-1082	8116978H1 (TONSDIC01) GNN g6114949_010.edit5p 6919538R8 (PLACFER06) GNN g6850654_000027_002 7368965H1 (ADREFEC01) 64460173H2 (OSTEUNC01) 6801172F6 (COLENOR03) 7212618T8 (LUNGFECE01) 6919538F8 (PLACFER06) 55073317H1 58003367H1 7271932R8 (OVARD101) 5623962R8 (THYMNOR02)	1 1497 1156 998 5742 5357 4290 3001 390 2592 4871 3542 4544	659 3728 1644 1496 6347 5883 4817 3712 1143 3387 5725 4220 5050

Table 4

Polynucleotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				72373545V1 5623962F8 (THYMNOR02)	1602 3970	2203 4319
33	5734965CB1	1876	1-902	3254961T6 (OWARTUN01)	1276	1876
				5897065H1 (BRAYDIN03)	1	291
				70810516V1	181	806
				70162895V1	1002	1658
				70809778V1	915	1490
				70807962V1	302	989
34	7473788CB1	1487	1-121, 1450-1487	70995937V1 7177378H1 (BRAXDIC01)	1024 29	1487 554
				GNN:g39833531_000002_00 2.edit.1	1	260
				70996158V1	594	1243
				7177563H2 (BRAXDIC01)	489	1180
35	3107989CB1	1884	1-306, 1253-1884	70942785V1 3107989F6 (BRSTTUT15)	1153 232	1507 609
				7363877H1 (OVARDIC01)	1358	1884
				GNN:g9368012.edit1	375	1465
				2243506F6 (PANCUT02)	1	385
36	7482887CB1	1070	1-660, 891-948	56009164H1 GBI:g5815507.edit GBI:g9716284_order_0.edit2	1 612	725 997
					988	1070
37	2963414CB1	2890	1-270, 1973-2064, 2658-2890, 726-1584	71883559V1 6741017F6 (BRAFDIT02) 72524920V1 7090654H1 (BRAUTDR03)	470 1687 984 2284	1087 2299 1725 2876

Table 4

Polymerotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				7595015H1 (L1VRNOC07)	1	450
				71882107V1	424	985
				70523289V1	1123	1749
				7236935H1 (BRAINNOY02)	1904	2302
				2601508H1 (UTRSNOT10)	2660	2890
38	7477139CB1	5198	2528-2698, 1296-2145, 2792-4455, 528-724, 177-214	GNN.g1149521_002 71143326V1 55117016H1 2879284F6 (UTRSTUT05) 3900926H1 (LUNGNON03)	948 4891 1 4388 3689	3957 5198 919 4874 3971
				GNN.g2780172_002.edit	3433	4943
				72615067V1	701	1315
				6775332H1 (OvardiR01)	4605	5193
				7369832H1 (ADREFEC01)	4063	4606
39	55009053CB1	3969	1393-2860, 1-649	8036923H1 (SMCRUNE01)	1289	2065
				72480126D1	3325	3969
				7263320F6 (PROSTM02)	1510	2343
				55009061H1	570	1318
				72476437D1	3306	3968
				6583144F8 (BRAVTEX01)	1	452
				72508467V1	2287	3200
				72509180V1	2494	3329
				55009045J1	288	982
40	7474648CB1	1803	198-1803	FL7474648_g7596812_0000	823	1497
				12_B7981277_1_1		
				GNN.g7596812_2	1	1803
41	7483053CB1	3472	1-305, 3134-3472	GB1.g6981824_000001.edit 2493520F6 (ADRETUT05)	1	337
					2055	2525

Table 4

Polymerucleotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				72498890V1	1524	2231
				GNN.g6981824_000001_042 74 edit	3187	
				55081239H1	847	1704
				6872245H1 (BRAGNONC02)	2354	3059
				7995993H1 (ADRETTUC01)	2942	3472
				7742567H1 (ADRETTUE04)	647	1183
42	7483117CB1	1704	1-342, 509-539, 582-758	GBL.g4153871_000001.edit	1536	1704
43	7484498CB1	6298	4050-4677, 1-195, 623-1785, 2406-2578, 3211-3637, 2139-2261	7369322F8 (ADREFEC01) GNN.g4153871_006.edit	343 1	501 1678
				55058386H1	601	1357
				7073440H1 (BRAUTDR04)	5165	5621
				7032228R8 (BRAXTDR12)	4000	4590
				55053104H1	1618	2321
				7014254F6 (KIDNNNOC01)	4579	5133
				7066070H1 (BRATNOR01)	2926	3470
				55053152H1	848	1564
				55058386J1	1	701
				7073642H1 (BRAUTDR04)	5045	5617
				6892089F6 (BRAITDR03)	2294	2708
				8267244H1 (MIXDUNF04)	4401	5097
				7076436H1 (BRAUTDR04)	3497	4047
				7068147R8 (BRATNOR01)	5186	5924
				GNN.g4508157_002.edit	1166	1941

Table 4

Polynucleotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				7741468H1 (THYMNOE01)	3001	3627
				6850478H1 (BRAIFFEN08)	5720	6298
				7068147F8 (BRAATNOR01)	4092	4592
44	7638121CB1	5454	1718-3145, 1-989, 3982-4016	675675311 (SINTFFER02), 736116H1 (BRAIFFE05), 55057003J1	3907	4637
				252	937	
				56000546J1	1303	2019
				7354408H1 (HEARNON03)	5008	5454
				5863411F6 (MUSLTD01)	3355	4178
				71873215V1	4520	5227
				71875134V1	3114	3669
				6496171T6 (COLNNOT41)	4710	5416
				55141853J2	810	1390
				7647137H1 (UTRSTUE01)	1920	2257
				7600017R6 (ESOGTME01)	1475	2041
				6200811F6 (PTTUNON01)	3037	3632
				55052669H1	2245	3081

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
24	7483046CB1	COLCTUT03
25	71636374CB1	CARDNOT01
26	7480597CB1	DRGLNOT01
27	3227248CB1	COTRNOT01
28	4207273CB1	TESTNOC01
29	7483334CB1	ADRENOT03
30	7483337CB1	UTRSTMR02
31	6035509CB1	PITUNOT06
32	7373485CB1	MCLDTXT02
33	5734965CB1	PROSTUS23
34	7473788CB1	BRAINOT19
35	3107989CB1	STOMFET02
37	2963414CB1	SCORNOT04
38	7477139CB1	PLACFER06
39	55009053CB1	SINITME01
41	7483053CB1	BRAYDIN03
42	7483117CB1	ADREFEC01
43	7484498CB1	BRAITDR03
44	7638121CB1	MUSLTDR02

Table 6

Library	Vector	Library Description
ADREFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from adrenal tissue removed from a Caucasian female fetus who died from anencephalus after 16-weeks' gestation. Serology was negative. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother.
ADREN0T03	PSPORT1	Library was constructed using RNA isolated from the adrenal tissue of a 17-year-old Caucasian male, who died from cerebral anoxia.
BRAIN0T19	pINCY	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology indicated a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. This tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. The left temporal lobe, including the mesial temporal structures, showed focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. GFAP was positive for astrocytes. The patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, and depressive disorder. Family history included brain cancer.
BRAITDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex, cingulate posterior tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAYDIN03	pINCY	This normalized library was constructed from 6.7 million independent clones from a brain tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48-hours/round) reannealing hybridization was used. The library was linearized and recircularized to select for insert containing clones.
CARDNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the cardiac muscle of a 65-year-old Caucasian male, who died from a gunshot wound.

Table 6

Library	Vector	Library Description
COLCTUT03	pINCY	Library was constructed using RNA isolated from cecal tumor tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology indicated invasive grade 2 adenocarcinoma forming an ulcerated mass 2 cm distal to the ileocecal valve and invading the muscularis propria. One regional lymph node (of 16) was positive for metastatic adenocarcinoma. Patient history included a deficiency anemia, malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, and normal delivery. Family history included cardiovascular and cerebrovascular disease, hyperlipidemia, and breast and ovarian cancer.
COTTRNOT01	pINCY	Library was constructed using RNA isolated from diseased transverse colon tissue obtained from a 26-year-old Caucasian male during a total abdominal colectomy and colostomy. Pathology indicated minimally active pancolitis with areas of focal severe colitis with perforation, consistent with Crohn's disease.
DRGILNOT01	pINCY	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus, infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
MCLDTXT02	pINCY	Library was constructed using RNA isolated from treated umbilical cord blood dendritic cells removed from a male. The cells were treated with granulocyte/macrophage colony stimulating factor (GM-CSF), tumor necrosis factor alpha (TNF alpha), stem cell factor (SCF), phorbol myristate acetate (PMA), and ionomycin. The GM-CSF was added at time 0 at 100 ng/ml, the TNF alpha was added at time 0 at 2.5 ng/ml, the SCF was added at time 0 at 25 ng/ml. The PMA and ionomycin were added at 13 days for five hours. Incubation time was 13 days.
MUSLTDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from right lower thigh muscle tissue removed from a 58-year-old Caucasian male during a wide resection of the right posterior thigh. Pathology indicated no residual tumor was identified in the right posterior thigh soft tissue. Changes were consistent with a previous biopsy site. On section through the soft tissue and muscle there was a smooth cystic cavity with hemorrhage around the margin on one side. The wall of the cyst was smooth and pale-tan. Pathology for the matched tumor tissue indicated a grade II liposarcoma. Patient history included liposarcoma (right thigh), and hypercholesterolemia. Previous surgeries included resection of right thigh mass. Family history included myocardial infarction and an unspecified rare blood disease.

Table 6

Library	Vector	Library Description
PITUNOT06	pINCY	Library was constructed using RNA isolated from pituitary gland tissue removed from a 55-year-old male who died from chronic obstructive pulmonary disease. Neuropathology indicated there were no gross abnormalities, other than mild ventricular enlargement. There was no apparent microscopic abnormality in any of the neocortical areas examined, except for a number of silver positive neurons with apical dendrite staining, particularly in the frontal lobe. The significance of this was undetermined. The only other microscopic abnormality was that there was prominent silver staining with some swollen axons in the CA3 region of the anterior and posterior hippocampus. Microscopic sections of the cerebellum revealed mild Bergmann's gliosis in the Purkinje cell layer. Patient history included schizophrenia.
PLACFER06	pINCY	This random primed library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.
PROSTUS23	pINCY	This subtracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate tissue.
SCORNTO4	pINCY	Library was constructed using RNA isolated from cervical spinal cord tissue removed from a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus, infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.

Table 6

Library	Vector	Library Description
SINTME01	pINCY	This 5' biased random primed library was constructed using RNA isolated from ileum tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the matched tumor tissue indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated 2 cm distal to the ileocecal valve. Patient history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, deficiency anemia, and normal delivery. Family history included breast cancer, atherosclerotic coronary artery disease, benign hypertension, cerebrovascular disease, ovarian cancer, and hyperlipidemia.
STOMFET02	pINCY	Library was constructed using RNA isolated from stomach tissue removed from a Hispanic male fetus, who died at 18 weeks' gestation.
TESTNOC01	PBLUESCRIPT	This large size fractionated library was constructed using RNA isolated from testicular tissue removed from a pool of eleven, 10 to 61-year-old Caucasian males.
UTRSTM02	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from endometrial tissue removed from a 32-year-old female (donor A) and using mRNA isolated from myometrium removed from a 45-year-old female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. In donor A, pathology indicated the endometrium was secretory phase. The cervix showed severe dysplasia (CIN III) focally involving the squamocolumnar junction at the 1, 6 and 7 o'clock positions. Mild koilocytotic dysplasia was also identified within the cervix. In donor B, pathology for the matched tumor tissue indicated multiple (23) subserosal, intramural, and submucosal leiomyomata. Patient history included stress incontinence, extrinsic asthma without status asthmaticus and normal delivery in donor B. Family history included cerebrovascular disease, depression, and atherosclerotic coronary artery disease in donor B.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-0; ESTs: Probability value= 1.0E-6; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value= 1.0E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value= 1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:320-322; Durbin, R. et al. (1998) <i>Our World View, in a Nutshell</i> , Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value=1.0E-3 or less; Signal peptide hits: Score= 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) <i>CABIOS</i> 4:61-66; Gribskov, M. et al. (1989) <i>Methods Enzymol.</i> 183:146-159; Bairoch, A. et al. (1997) <i>Nucleic Acids Res.</i> 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) <i>Genome Res.</i> 8:175-185; Ewing, B. and P. Green (1998) <i>Genome Res.</i> 8:186-194.	
Phrap	A Phred Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489; Smith, T.F. and M.S. Waterman (1981) <i>J. Mol. Biol.</i> 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) <i>Genome Res.</i> 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) <i>Protein Engineering</i> 10:1-6; Claverie, J.M. and S. Audic (1997) <i>CABIOS</i> 12:431-439.	Score=3.5 or greater

**Table 7**

Program	Description	Reference	Parameter Threshold
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Person, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Person, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Somhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and
  - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

5 10. A method of claim 9, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

10 12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).

20 13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

25 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

30 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 5 a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

10

18. A composition of claim 17, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

19. A method for treating a disease or condition associated with decreased expression of 15 functional PKIN, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- 20 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

25

22. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 21.

30 23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional  
5 PKIN, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- 10 a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of  
15 claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 20 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

25  
28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 30 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

10 30. A diagnostic test for a condition or disease associated with the expression of PKIN in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

20 25 31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

30

33. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

5

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

10

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

20

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

25

30

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

5

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

10 44. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ 15 ID NO:1-22 in the sample.

45. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

20

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 25 13.

25

47. A method of generating a transcript image of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

30

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

5

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is 10 completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

15 52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

20 54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

25 55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

30 56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

5 61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

10

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

15

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

20

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

25

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

30

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.
- 5 78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.
79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.
80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.
- 10 81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.
82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.
- 15 83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.
84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.
85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.
- 20 86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.
87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.
- 25 88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.
89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.
90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.
- 30 91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.
92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

5 95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

10

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

<110> INCYTE GENOMICS, INC.  
GURURAJAN, Rajagopal  
BAUGHN, Mariah R.  
WALIA, Narinder K.  
ELLIOTT, Vicki S.  
XU, Yuming  
ARVIZU, Chandra  
YAO, Monique G.  
RAMKUMAR, Jayalaxmi  
DING, Li  
TANG, Y. Tom  
HAFALIA, April J.A.  
NGUYEN, Dannie B.  
GANDHI, Ameena R.  
LU, Yan  
YUE, Henry  
BURFORD, Neil  
BANDMAN, Olga  
TRIBOULEY, Catherine  
LAL, Preeti G.  
RECIPON, Shirley A.  
LU, Dyung Aina M.  
BOROWSKY, Mark L.  
THORNTON, Michael  
SWARNAKER Anita  
THANGAVELU, Kavitha  
KHAN, Farrah A.  
ISON, Craig H.

<120> HUMAN KINASES

<130> PI-0262 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/242,410; 60/244,068; 60/245,708; 60/247,672; 60/249,565;  
60/252,730; 60/250,807

<151> 2000-10-20; 2000-10-27; 2000-11-03; 2000-11-09; 2000-11-16  
2000-11-22; 2000-12-01

<160> 44

<170> PERL Program

<210> 1

<211> 337

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7482896CD1

<400> 1

Met Thr Asn Asn Ser Gly Ser Lys Ala Glu Leu Val Val Gly Gly  
1 5 10 15

Val Tyr Leu Gly Ile Thr Thr Thr Asn Gly Glu Asp Val Ala Val  
                   35                  40                  45  
 Lys Leu Glu Ser Gln Lys Val Lys His Pro Gln Leu Leu Tyr Glu  
                   50                  55                  60  
 Ser Lys Leu Tyr Thr Ile Leu Gln Gly Gly Val Gly Ile Pro His  
                   65                  70                  75  
 Met His Trp Tyr Gly Gln Glu Lys Asp Asn Asn Val Leu Val Met  
                   80                  85                  90  
 Asp Leu Leu Gly Pro Ser Leu Glu Asp Leu Phe Asn Phe Cys Ser  
                   95                  100                105  
 Arg Arg Phe Thr Met Lys Thr Val Leu Met Leu Ala Asp Gln Met  
                   110                115                120  
 Ile Ser Arg Ile Glu Tyr Val His Thr Lys Asn Phe Leu His Arg  
                   125                130                135  
 Asp Ile Lys Pro Asp Asn Phe Leu Met Gly Thr Gly Arg His Cys  
                   140                145                150  
 Asn Lys Leu Phe Leu Ile Asp Phe Gly Leu Ala Lys Lys Tyr Arg  
                   155                160                165  
 Asp Asn Arg Thr Arg Gln His Ile Pro Tyr Arg Glu Asp Lys His  
                   170                175                180  
 Leu Ile Gly Thr Val Arg Tyr Ala Ser Ile Asn Ala His Leu Gly  
                   185                190                195  
 Ile Glu Gln Ser Arg Arg Asp Asp Met Glu Ser Leu Gly Tyr Val  
                   200                205                210  
 Phe Met Tyr Phe Asn Arg Thr Ser Leu Pro Trp Gln Gly Leu Arg  
                   215                220                225  
 Ala Met Thr Lys Lys Gln Lys Tyr Glu Lys Ile Ser Glu Lys Lys  
                   230                235                240  
 Met Ser Thr Pro Val Glu Val Leu Cys Lys Gly Phe Pro Ala Glu  
                   245                250                255  
 Phe Ala Met Tyr Leu Asn Tyr Cys Arg Gly Leu Arg Phe Glu Glu  
                   260                265                270  
 Val Pro Asp Tyr Met Tyr Leu Arg Gln Leu Phe Arg Ile Leu Phe  
                   275                280                285  
 Arg Thr Leu Asn His Gln Tyr Asp Tyr Thr Phe Asp Trp Thr Met  
                   290                295                300  
 Leu Lys Gln Lys Ala Ala Gln Gln Ala Ala Ser Ser Ser Gly Gln  
                   305                310                315  
 Gly Gln Gln Ala Gln Thr Gln Thr Gly Lys Gln Thr Glu Lys Asn  
                   320                325                330  
 Lys Asn Asn Val Lys Asp Asn  
                   335

<210> 2  
 <211> 475  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7483046CD1

<400> 2  
 Met Glu Arg Arg Leu Arg Ala Leu Glu Gln Leu Ala Arg Gly Glu  
   1                  5                  10                  15  
 Ala Gly Gly Cys Pro Gly Leu Asp Gly Leu Leu Asp Leu Leu

20	25	30
Ala Leu His His Glu	Leu Ser Ser Gly	Pro Leu Arg Arg Glu Arg
35	40	45
Ser Val Ala Gln Phe	Leu Ser Trp Ala Ser	Pro Phe Val Ser Lys
50	55	60
Val Lys Glu Leu Arg	Leu Gln Arg Asp Asp	Phe Glu Ile Leu Lys
65	70	75
Val Ile Gly Arg Gly	Ala Phe Gly Glu Val	Thr Val Val Arg Gln
80	85	90
Arg Asp Thr Gly Gln	Ile Phe Ala Met Lys	Met Leu His Lys Trp
95	100	105
Glu Met Leu Lys Arg	Ala Glu Thr Ala Cys	Phe Arg Glu Glu Arg
110	115	120
Asp Val Leu Val Lys	Gly Asp Ser Arg	Trp Val Thr Thr Leu His
125	130	135
Tyr Ala Phe Gln Asp	Glu Glu Tyr Leu Tyr	Leu Val Met Asp Tyr
140	145	150
Tyr Ala Gly Gly Asp	Leu Leu Thr Leu	Leu Ser Arg Phe Glu Asp
155	160	165
Arg Leu Pro Pro Glu	Leu Ala Gln Phe	Tyr Leu Ala Glu Met Val
170	175	180
Leu Ala Ile His Ser	Leu His Gln Leu	Gly Tyr Val His Arg Asp
185	190	195
Val Lys Pro Asp Asn	Val Leu Leu Asp	Val Asn Gly His Ile Arg
200	205	210
Leu Ala Asp Phe Gly	Ser Cys Leu Arg	Leu Asn Thr Asn Gly Met
215	220	225
Val Asp Ser Ser Val	Ala Val Gly Thr	Pro Asp Tyr Ile Ser Pro
230	235	240
Glu Ile Leu Gln Ala	Met Glu Glu Gly	Lys Gly His Tyr Gly Pro
245	250	255
Gln Cys Asp Trp	Trp Ser Leu Gly Val	Cys Ala Tyr Glu Leu Leu
260	265	270
Phe Gly Glu Thr Pro	Phe Tyr Ala Glu	Ser Leu Val Glu Thr Tyr
275	280	285
Gly Lys Ile Met Asn	His Glu Asp His	Leu Gln Phe Pro Pro Asp
290	295	300
Val Pro Asp Val Pro	Ala Ser Ala Gln	Asp Leu Ile Arg Gln Leu
305	310	315
Leu Cys Arg Gln	Glu Glu Arg Leu Gly	Arg Gly Leu Asp Asp
320	325	330
Phe Arg Asn His Pro	Phe Phe Glu Gly	Val Asp Trp Glu Arg Leu
335	340	345
Ala Ser Ser Thr Ala	Pro Tyr Ile Pro	Glu Leu Arg Gly Pro Met
350	355	360
Asp Thr Ser Asn Phe	Asp Val Asp Asp	Asp Thr Leu Asn His Pro
365	370	375
Gly Thr Leu Pro Pro	Pro Ser His Gly	Ala Phe Ser Gly His His
380	385	390
Leu Pro Phe Val	Gly Phe Thr Tyr	Thr Ser Gly Ser His Ser Pro
395	400	405
Glu Ser Ser Ser Glu	Ala Trp Ala Ala	Leu Glu Arg Lys Leu Gln
410	415	420
Cys Leu Glu Gln	Glu Lys Val Glu Leu	Ser Arg Lys His Gln Glu
425	430	435
Ala Leu His Ala Pro	Thr Asp His Arg	Glu Leu Glu Gln Leu Arg

440	445	450
Lys Glu Val Gln Thr Leu Arg Asp Arg	Leu Pro Gly Ile Pro Ser	
455	460	465
Ala His Pro His Pro Leu Leu Glu Phe Leu		
470	475	

<210> 3  
 <211> 675  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 71636374CD1

<400> 3		
Met Thr Thr Ser His Met Asn Gly His Val Thr Glu Glu Ser Asp		
1 5 10 15		
Ser Glu Val Lys Asn Val Asp Leu Ala Ser Pro Glu Glu His Gln		
20 25 30		
Lys His Arg Glu Met Ala Val Asp Cys Pro Gly Asp Leu Gly Thr		
35 40 45		
Arg Met Met Pro Ile Arg Arg Ser Ala Gln Leu Glu Arg Ile Arg		
50 55 60		
Gln Gln Glu Asp Met Arg Arg Arg Glu Glu Glu Gly Lys		
65 70 75		
Lys Gln Glu Leu Asp Leu Asn Ser Ser Met Arg Leu Lys Lys Leu		
80 85 90		
Ala Gln Ile Pro Pro Lys Thr Gly Ile Asp Asn Pro Met Phe Asp		
95 100 105		
Thr Glu Glu Gly Ile Val Leu Glu Ser Pro His Tyr Ala Val Lys		
110 115 120		
Ile Leu Glu Ile Glu Asp Leu Phe Ser Ser Leu Lys His Ile Gln		
125 130 135		
His Thr Leu Val Asp Ser Gln Ser Gln Glu Asp Ile Ser Leu Leu		
140 145 150		
Leu Gln Leu Val Gln Asn Lys Asp Phe Gln Asn Ala Phe Lys Ile		
155 160 165		
His Asn Ala Ile Thr Val His Met Asn Lys Ala Ser Pro Pro Phe		
170 175 180		
Pro Leu Ile Ser Asn Ala Gln Asp Leu Ala Gln Glu Val Gln Thr		
185 190 195		
Val Leu Lys Pro Val His His Lys Glu Gly Gln Glu Leu Thr Ala		
200 205 210		
Leu Leu Asn Thr Pro His Ile Gln Ala Leu Leu Leu Ala His Asp		
215 220 225		
Lys Val Ala Glu Gln Glu Met Gln Leu Glu Pro Ile Thr Asp Glu		
230 235 240		
Arg Val Tyr Glu Ser Ile Gly Gln Tyr Gly Gly Glu Thr Val Lys		
245 250 255		
Ile Val Arg Ile Glu Lys Ala Arg Asp Ile Pro Leu Gly Ala Thr		
260 265 270		
Val Arg Asn Glu Met Asp Ser Val Ile Ile Ser Arg Ile Val Lys		
275 280 285		
Gly Gly Ala Ala Glu Lys Ser Gly Leu Leu His Glu Gly Asp Glu		
290 295 300		

Val Leu Glu Ile Asn Gly Ile Glu Ile Arg Gly Lys Asp Val Asn  
305 310 315  
Glu Val Phe Asp Leu Leu Ser Asp Met His Gly Thr Leu Thr Phe  
320 325 330  
Val Leu Ile Pro Ser Gln Gln Ile Lys Pro Pro Pro Ala Lys Glu  
335 340 345  
Thr Val Ile His Val Lys Ala His Phe Asp Tyr Asp Pro Ser Asp  
350 355 360  
Asp Pro Tyr Val Pro Cys Arg Glu Leu Gly Leu Ser Phe Gln Lys  
365 370 375  
Gly Asp Ile Leu His Val Ile Ser Gln Glu Asp Pro Asn Trp Trp  
380 385 390  
Gln Ala Tyr Arg Glu Gly Asp Glu Asp Asn Gln Pro Leu Ala Gly  
395 400 405  
Leu Val Pro Gly Lys Ser Phe Gln Gln Arg Glu Ala Met Lys  
410 415 420  
Gln Thr Ile Glu Glu Asp Lys Glu Pro Glu Lys Ser Gly Lys Leu  
425 430 435  
Trp Cys Ala Lys Lys Asn Lys Lys Arg Lys Lys Val Leu Tyr  
440 445 450  
Asn Ala Asn Lys Asn Asp Asp Tyr Asp Asn Glu Glu Ile Leu Thr  
455 460 465  
Tyr Glu Glu Met Ser Leu Tyr His Gln Pro Ala Asn Arg Lys Arg  
470 475 480  
Pro Ile Ile Leu Ile Gly Pro Gln Asn Cys Gly Gln Asn Glu Leu  
485 490 495  
Arg Gln Arg Leu Met Asn Lys Glu Lys Asp Arg Phe Ala Ser Ala  
500 505 510  
Val Pro His Thr Thr Arg Ser Arg Arg Asp Gln Glu Val Ala Gly  
515 520 525  
Arg Asp Tyr His Phe Val Ser Arg Gln Ala Phe Glu Ala Asp Ile  
530 535 540  
Ala Ala Gly Lys Phe Ile Glu His Gly Glu Phe Glu Lys Asn Leu  
545 550 555  
Tyr Gly Thr Ser Ile Asp Ser Val Arg Gln Val Ile Asn Ser Gly  
560 565 570  
Lys Ile Cys Leu Leu Ser Leu Arg Thr Gln Ser Leu Lys Thr Leu  
575 580 585  
Arg Asn Ser Asp Leu Lys Pro Tyr Ile Ile Phe Ile Ala Pro Pro  
590 595 600  
Ser Gln Glu Arg Leu Arg Ala Leu Leu Ala Lys Glu Gly Lys Asn  
605 610 615  
Pro Lys Pro Glu Glu Leu Arg Glu Ile Ile Glu Lys Thr Arg Glu  
620 625 630  
Met Glu Gln Asn Asn Gly His Tyr Phe Asp Thr Ala Ile Val Asn  
635 640 645  
Ser Asp Leu Asp Lys Ala Tyr Gln Glu Leu Leu Arg Leu Ile Asn  
650 655 660  
Lys Leu Asp Thr Glu Pro Gln Trp Val Pro Ser Thr Trp Leu Arg  
665 670 675

<210> 4  
<211> 835  
<212> PRT  
<213> Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7480597CD1

&lt;400&gt; 4

Met	Ala	Glu	Gly	Lys	Glu	Gly	Gln	Val	Pro	Ser	Tyr	Met	Asp	Gly
1				5				10				15		
Ser	Arg	Gln	Arg	Glu	Asn	Glu	Glu	Asp	Ala	Lys	Ala	Glu	Thr	Pro
				20				25				30		
Asp	Val	Thr	Ile	Arg	Ser	Tyr	Glu	Ile	Tyr	Ser	Leu	Pro	Trp	Asn
				35				40				45		
Arg	Gln	Gln	Gly	Leu	Cys	Asp	His	Ser	Leu	Lys	Tyr	Leu	Ser	Ser
				50				55				60		
Arg	Ile	Thr	Glu	Arg	Lys	Leu	Gln	Gly	Ser	Trp	Leu	Pro	Ala	Ser
				65				70				75		
Arg	Gly	Asn	Leu	Glu	Lys	Pro	Phe	Leu	Gly	Pro	Arg	Gly	Pro	Val
				80				85				90		
Val	Pro	Leu	Phe	Cys	Pro	Arg	Asn	Gly	Leu	His	Ser	Ala	His	Pro
				95				100				105		
Glu	Asn	Ser	Pro	Leu	Lys	Pro	Arg	Val	Val	Thr	Val	Val	Lys	Leu
				110				115				120		
Gly	Gly	Gln	Arg	Pro	Arg	Lys	Ile	Thr	Leu	Leu	Leu	Asn	Arg	Arg
				125				130				135		
Ser	Val	Gln	Thr	Phe	Glu	Gln	Leu	Leu	Ala	Asp	Ile	Ser	Glu	Ala
				140				145				150		
Leu	Gly	Ser	Pro	Arg	Trp	Lys	Asn	Asp	Arg	Val	Arg	Lys	Leu	Phe
				155				160				165		
Asn	Leu	Lys	Gly	Arg	Glu	Ile	Arg	Ser	Val	Ser	Asp	Phe	Phe	Arg
				170				175				180		
Glu	Gly	Asp	Ala	Phe	Ile	Ala	Met	Gly	Lys	Glu	Pro	Leu	Thr	Leu
				185				190				195		
Lys	Ser	Ile	Gln	Val	Ala	Val	Glu	Glu	Leu	Tyr	Pro	Asn	Lys	Ala
				200				205				210		
Arg	Ala	Leu	Thr	Leu	Ala	Gln	His	Ser	Arg	Ala	Pro	Ser	Pro	Arg
				215				220				225		
Leu	Arg	Ser	Arg	Leu	Phe	Ser	Lys	Ala	Leu	Lys	Gly	Asp	His	Arg
				230				235				240		
Cys	Gly	Glu	Thr	Glu	Thr	Pro	Lys	Ser	Cys	Ser	Glu	Val	Ala	Gly
				245				250				255		
Cys	Lys	Ala	Ala	Met	Arg	His	Gln	Gly	Lys	Ile	Pro	Glu	Glu	Leu
				260				265				270		
Ser	Leu	Asp	Asp	Arg	Ala	Arg	Thr	Gln	Lys	Lys	Trp	Gly	Arg	Gly
				275				280				285		
Lys	Trp	Glu	Pro	Glu	Pro	Ser	Ser	Lys	Pro	Pro	Arg	Glu	Ala	Thr
				290				295				300		
Leu	Glu	Glu	Arg	His	Ala	Arg	Gly	Glu	Lys	His	Leu	Gly	Val	Glu
				305				310				315		
Ile	Glu	Lys	Thr	Ser	Gly	Glu	Ile	Ile	Arg	Cys	Glu	Lys	Cys	Lys
				320				325				330		
Arg	Glu	Arg	Glu	Leu	Gln	Gln	Ser	Leu	Glu	Arg	Glu	Arg	Leu	Ser
				335				340				345		
Leu	Gly	Thr	Ser	Glu	Leu	Asp	Met	Gly	Lys	Gly	Pro	Met	Tyr	Asp
				350				355				360		
Val	Glu	Lys	Leu	Val	Arg	Thr	Arg	Ser	Cys	Arg	Arg	Ser	Pro	Glu
				365				370				375		
Ala	Asn	Pro	Ala	Ser	Gly	Glu	Glu	Gly	Trp	Lys	Gly	Asp	Ser	His

380	385	390
Arg Ser Ser Pro Arg Asn Pro Thr Gln	Glu Leu Arg Arg Pro Ser	
395	400	405
Lys Ser Met Asp Lys Lys Glu Asp Arg Gly Pro Glu Asp Gln Glu		
410	415	420
Ser His Ala Gln Gly Ala Ala Lys Ala Lys Lys Asp Leu Val Glu		
425	430	435
Val Leu Pro Val Thr Glu Glu Gly Leu Arg Glu Val Lys Lys Asp		
440	445	450
Thr Arg Pro Met Ser Arg Ser Lys His Gly Gly Trp Leu Leu Arg		
455	460	465
Glu His Gln Ala Gly Phe Glu Lys Leu Arg Arg Thr Arg Gly Glu		
470	475	480
Glu Lys Glu Ala Glu Lys Glu Lys Pro Cys Met Ser Gly Gly		
485	490	495
Arg Arg Met Thr Leu Arg Asp Asp Gln Pro Ala Lys Leu Glu Lys		
500	505	510
Glu Pro Lys Thr Arg Pro Glu Glu Asn Lys Pro Glu Arg Pro Ser		
515	520	525
Gly Arg Lys Pro Arg Pro Met Gly Ile Ile Ala Ala Asn Val Glu		
530	535	540
Lys His Tyr Glu Thr Gly Arg Val Ile Gly Asp Gly Asn Phe Ala		
545	550	555
Val Val Lys Glu Cys Arg His Arg Glu Thr Arg Gln Ala Tyr Ala		
560	565	570
Met Lys Ile Ile Asp Lys Ser Arg Leu Lys Gly Lys Glu Asp Met		
575	580	585
Val Asp Ser Glu Ile Leu Ile Ile Gln Ser Leu Ser His Pro Asn		
590	595	600
Ile Val Lys Leu His Glu Val Tyr Glu Thr Asp Met Glu Ile Tyr		
605	610	615
Leu Ile Leu Glu Tyr Val Gln Gly Gly Asp Leu Phe Asp Ala Ile		
620	625	630
Ile Glu Ser Val Lys Phe Pro Glu Pro Asp Ala Ala Leu Met Ile		
635	640	645
Met Asp Leu Cys Lys Ala Leu Val His Met His Asp Lys Ser Ile		
650	655	660
Val His Arg Asp Leu Lys Pro Glu Asn Leu Leu Val Gln Arg Asn		
665	670	675
Glu Asp Lys Ser Thr Thr Leu Lys Leu Ala Asp Phe Gly Leu Ala		
680	685	690
Lys His Val Val Arg Pro Ile Phe Thr Val Cys Gly Thr Pro Thr		
695	700	705
Tyr Val Ala Pro Glu Ile Leu Ser Glu Lys Gly Tyr Gly Leu Glu		
710	715	720
Val Asp Met Trp Ala Ala Gly Val Ile Leu Tyr Ile Leu Leu Cys		
725	730	735
Gly Phe Pro Pro Phe Arg Ser Pro Glu Arg Asp Gln Asp Glu Leu		
740	745	750
Phe Asn Ile Ile Gln Leu Gly His Phe Glu Phe Leu Pro Pro Tyr		
755	760	765
Trp Asp Asn Ile Ser Asp Ala Ala Lys Asp Leu Val Ser Arg Leu		
770	775	780
Leu Val Val Asp Pro Lys Lys Arg Tyr Thr Ala His Gln Val Leu		
785	790	795
Gln His Pro Trp Ile Glu Thr Ala Gly Lys Thr Asn Thr Val Lys		

800	805	810
Arg Gln Lys Gln Val Ser Pro Ser Ser	Glu Gly His Phe Arg Ser	
815	820	825
Gln His Lys Arg Val Val Glu Gln Val Ser		
830	835	

<210> 5

<211> 373

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3227248CD1

<400> 5

Met Lys Leu Ile Asn Gly Lys Lys Gln Thr Phe Pro Trp Phe Gly			
1	5	10	15
Met Asp Ile Gly Gly Thr Leu Val Lys Leu Val Tyr Phe Glu Pro			
20	25	30	
Lys Asp Ile Thr Ala Glu Glu Glu Gln Glu Glu Val Glu Asn Leu			
35	40	45	
Lys Ser Ile Arg Lys Tyr Leu Thr Ser Asn Thr Ala Tyr Gly Lys			
50	55	60	
Thr Gly Ile Arg Asp Val His Leu Glu Leu Lys Asn Leu Thr Met			
65	70	75	
Cys Gly Arg Lys Gly Asn Leu His Phe Ile Arg Phe Pro Ser Cys			
80	85	90	
Ala Met His Arg Phe Ile Gln Met Gly Ser Glu Lys Asn Phe Ser			
95	100	105	
Ser Leu His Thr Thr Leu Cys Ala Thr Gly Gly Ala Phe Lys			
110	115	120	
Phe Glu Glu Asp Phe Arg Met Ile Ala Asp Leu Gln Leu His Lys			
125	130	135	
Leu Asp Glu Leu Asp Cys Leu Ile Gln Gly Leu Leu Tyr Val Asp			
140	145	150	
Ser Val Gly Phe Asn Gly Lys Pro Glu Cys Tyr Tyr Phe Glu Asn			
155	160	165	
Pro Thr Asn Pro Glu Leu Cys Gln Lys Lys Pro Tyr Cys Leu Asp			
170	175	180	
Asn Pro Tyr Pro Met Leu Leu Val Asn Met Gly Ser Gly Val Ser			
185	190	195	
Ile Leu Ala Val Tyr Ser Lys Asp Asn Tyr Lys Arg Val Thr Gly			
200	205	210	
Thr Ser Leu Gly Gly Thr Phe Leu Gly Leu Cys Cys Leu Leu			
215	220	225	
Thr Gly Cys Glu Thr Phe Glu Glu Ala Leu Glu Met Ala Ala Lys			
230	235	240	
Gly Asp Ser Thr Asn Val Asp Lys Leu Val Lys Asp Ile Tyr Gly			
245	250	255	
Gly Asp Tyr Glu Arg Phe Gly Leu Gln Gly Ser Ala Val Ala Ser			
260	265	270	
Ser Phe Gly Asn Met Met Ser Lys Glu Lys Arg Asp Ser Ile Ser			
275	280	285	
Lys Glu Asp Leu Ala Arg Ala Thr Leu Val Thr Ile Thr Asn Asn			
290	295	300	

Ile Gly Ser Ile Ala Arg Met Cys Ala Leu Asn Glu Asn Ile Asp  
 305 310 315  
 Arg Val Val Phe Val Gly Asn Phe Leu Arg Ile Asn Met Val Ser  
 320 325 330  
 Met Lys Leu Leu Ala Tyr Ala Met Asp Phe Trp Ser Lys Gly Gln  
 335 340 345  
 Leu Lys Ala Leu Phe Leu Glu His Glu Gly Tyr Phe Gly Ala Val  
 350 355 360  
 Gly Ala Leu Leu Glu Leu Phe Lys Met Thr Asp Asp Lys  
 365 370

&lt;210&gt; 6

&lt;211&gt; 735

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4207273CD1

&lt;400&gt; 6

Met Pro Gln Ile Ala Lys Lys Gln Ser Thr His Arg Thr Gln Lys  
 1 5 10 15  
 Pro Lys Lys Gln Ser Phe Pro Cys Ile Cys Lys Asn Pro Gly Thr  
 20 25 30  
 Gln Lys Ser Cys Val Pro Leu Ser Val Gln Pro Thr Glu Pro Arg  
 35 40 45  
 Leu Asn Tyr Leu Asp Leu Lys Tyr Ser Asp Met Phe Lys Glu Ile  
 50 55 60  
 Asn Ser Thr Ala Asn Gly Pro Gly Ile Tyr Glu Met Phe Gly Thr  
 65 70 75  
 Pro Val Tyr Cys His Val Arg Glu Thr Glu Arg Asp Glu Asn Thr  
 80 85 90  
 Tyr Tyr Arg Glu Ile Cys Ser Ala Pro Ser Gly Arg Arg Ile Thr  
 95 100 105  
 Asn Lys Cys Arg Ser Ser His Ser Glu Arg Lys Ser Asn Ile Arg  
 110 115 120  
 Thr Arg Leu Ser Gln Lys Lys Thr His Met Lys Cys Pro Lys Thr  
 125 130 135  
 Ser Phe Gly Ile Lys Gln Glu His Lys Val Leu Ile Ser Lys Glu  
 140 145 150  
 Lys Ser Ser Lys Ala Val His Ser Asn Leu His Asp Ile Glu Asn  
 155 160 165  
 Gly Asp Gly Ile Ser Glu Pro Asp Trp Gln Ile Lys Ser Ser Gly  
 170 175 180  
 Asn Glu Phe Leu Ser Ser Lys Asp Glu Ile His Pro Met Asn Leu  
 185 190 195  
 Ala Gln Thr Pro Glu Gln Ser Met Lys Gln Asn Glu Phe Pro Pro  
 200 205 210  
 Val Ser Asp Leu Ser Ile Val Glu Glu Val Ser Met Glu Glu Ser  
 215 220 225  
 Thr Gly Asp Arg Asp Ile Ser Asn Asn Gln Ile Leu Thr Thr Ser  
 230 235 240  
 Leu Arg Asp Leu Gln Glu Leu Glu Glu Leu His His Gln Ile Pro  
 245 250 255  
 Phe Ile Pro Ser Glu Asp Ser Trp Ala Val Pro Ser Glu Lys Asn

260	265	270
Ser Asn Lys Tyr Val Gln Gln Glu Lys	Gln Asn Thr Ala Ser	Leu
275	280	285
Ser Lys Val Asn Ala Ser Arg Ile Leu	Thr Asn Asp Leu Glu	Phe
290	295	300
Asp Ser Val Ser Asp His Ser Lys Thr	Leu Thr Asn Phe Ser	Phe
305	310	315
Gln Ala Lys Gln Glu Ser Ala Ser Ser	Gln Thr Tyr Gln Tyr	Trp
320	325	330
Val His Tyr Leu Asp His Asp Ser Leu	Ala Asn Lys Ser Ile	Thr
335	340	345
Tyr Gln Met Phe Gly Lys Thr Leu Ser	Gly Thr Asn Ser Ile	Ser
350	355	360
Gln Glu Ile Met Asp Ser Val Asn Asn	Glu Glu Leu Thr Asp	Glu
365	370	375
Leu Leu Gly Cys Leu Ala Ala Glu Leu	Leu Ala Leu Asp Glu	Lys
380	385	390
Asp Asn Asn Ser Cys Gln Lys Met Ala	Asn Glu Thr Asp Pro	Glu
395	400	405
Asn Leu Asn Leu Val Leu Arg Trp Arg	Gly Ser Thr Pro	Lys Glu
410	415	420
Met Gly Arg Glu Thr Thr Lys Val Lys	Ile Gln Arg His Ser	Ser
425	430	435
Gly Leu Arg Ile Tyr Asp Arg Glu Glu	Lys Phe Leu Ile Ser	Asn
440	445	450
Glu Lys Lys Ile Phe Ser Glu Asn Ser	Leu Lys Ser Glu Glu	Pro
455	460	465
Ile Leu Trp Thr Lys Gly Glu Ile Leu	Gly Lys Gly Ala Tyr	Gly
470	475	480
Thr Val Tyr Cys Gly Leu Thr Ser Gln	Gly Gln Leu Ile Ala	Val
485	490	495
Lys Gln Val Ala Leu Asp Thr Ser Asn	Lys Leu Ala Ala Glu	Lys
500	505	510
Glu Tyr Arg Lys Leu Gln Glu Glu Val	Asp Leu Leu Lys Ala	Leu
515	520	525
Lys His Val Asn Ile Val Ala Tyr Leu	Gly Thr Cys Leu Gln	Glu
530	535	540
Asn Thr Val Ser Ile Phe Met Glu Phe	Val Pro Gly Gly Ser	Ile
545	550	555
Ser Ser Ile Ile Asn Arg Phe Gly Pro	Leu Pro Glu Met Val	Phe
560	565	570
Cys Lys Tyr Thr Lys Gln Ile Leu Gln	Gly Val Ala Tyr Leu	His
575	580	585
Glu Asn Cys Val Val His Arg Asp Ile	Lys Gly Asn Asn Val	Met
590	595	600
Leu Met Pro Thr Gly Ile Ile Lys Leu	Ile Asp Phe Gly Cys	Ala
605	610	615
Arg Arg Leu Ala Trp Ala Gly Leu Asn	Gly Thr His Ser Asp	Met
620	625	630
Leu Lys Ser Met His Gly Thr Pro Tyr	Trp Met Ala Pro Glu	Val
635	640	645
Ile Asn Glu Ser Gly Tyr Gly Arg Lys	Ser Asp Ile Trp Ser	Ile
650	655	660
Gly Cys Thr Val Phe Glu Met Ala Thr	Gly Lys Pro Pro Leu	Ala
665	670	675
Ser Met Asp Arg Met Ala Ala Met Phe	Tyr Ile Gly Ala His	Arg

680	685	690
Gly Leu Met Pro Pro Leu Pro Asp His	Phe Ser Glu Asn Ala Ala	
695	700	705
Asp Phe Val Arg Met Cys Leu Thr Arg	Asp Gln His Glu Arg Pro	
710	715	720
Ser Ala Leu Gln Leu Leu Lys His Ser	Phe Leu Glu Arg Ser His	
725	730	735

<210> 7  
 <211> 506  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7483334CD1

<400> 7

Met Asp Asp Tyr Met Val Leu Arg Met Ile Gly Glu Gly Ser Phe			
1	5	10	15
Gly Arg Ala Leu Leu Val Gln Leu Glu Ser Ser Asn Gln Met Phe			
20	25	30	
Ala Met Lys Glu Ile Arg Leu Pro Lys Ser Phe Ser Asn Thr Gln			
35	40	45	
Asn Ser Arg Lys Glu Ala Val Leu Leu Ala Lys Met Lys His Pro			
50	55	60	
Asn Ile Val Ala Phe Lys Glu Ser Phe Glu Ala Glu Gly His Leu			
65	70	75	
Tyr Ile Val Met Glu Tyr Cys Asp Gly Gly Asp Leu Met Gln Lys			
80	85	90	
Ile Lys Gln Gln Lys Gly Lys Leu Phe Pro Glu Asp Met Ile Leu			
95	100	105	
Asn Trp Phe Thr Gln Met Cys Leu Gly Val Asn His Ile His Lys			
110	115	120	
Lys Arg Val Leu His Arg Asp Ile Lys Ser Lys Asn Ile Phe Leu			
125	130	135	
Thr Gln Asn Gly Lys Val Lys Leu Gly Asp Phe Gly Ser Ala Arg			
140	145	150	
Leu Leu Ser Asn Pro Met Ala Phe Ala Cys Thr Tyr Val Gly Thr			
155	160	165	
Pro Tyr Tyr Val Pro Pro Glu Ile Trp Glu Asn Leu Pro Tyr Asn			
170	175	180	
Asn Lys Ser Asp Ile Trp Ser Leu Gly Cys Ile Leu Tyr Glu Leu			
185	190	195	
Cys Thr Leu Lys His Pro Phe Gln Ala Asn Ser Trp Lys Asn Leu			
200	205	210	
Ile Leu Lys Val Cys Gln Gly Cys Ile Ser Pro Leu Pro Ser His			
215	220	225	
Tyr Ser Tyr Glu Leu Gln Phe Leu Val Lys Gln Met Phe Lys Arg			
230	235	240	
Asn Pro Ser His Arg Pro Ser Ala Thr Thr Leu Leu Ser Arg Gly			
245	250	255	
Ile Val Ala Arg Leu Val Gln Lys Cys Leu Pro Pro Glu Ile Ile			
260	265	270	
Met Glu Tyr Gly Glu Glu Val Leu Glu Glu Ile Lys Asn Ser Lys			

275	280	285
His Asn Thr Pro Arg Lys Lys Thr Asn	Pro Ser Arg Ile Arg	Ile
290	295	300
Ala Leu Gly Asn Glu Ala Ser Thr Val	Gln Glu Glu Glu Gln Asp	
305	310	315
Arg Lys Gly Ser His Thr Asp Leu Glu	Ser Ile Asn Glu Asn Leu	
320	325	330
Val Glu Ser Ala Leu Arg Arg Val Asn	Arg Glu Glu Lys Gly Asn	
335	340	345
Lys Ser Val His Leu Arg Lys Ala Ser	Ser Pro Asn Leu His Arg	
350	355	360
Arg Gln Trp Glu Lys Asn Val Pro Asn	Thr Ala Leu Thr Ala Leu	
365	370	375
Glu Asn Ala Ser Ile Leu Thr Ser Ser	Leu Thr Ala Glu Asp Asp	
380	385	390
Arg Gly Gly Ser Val Ile Lys Tyr Ser	Lys Asn Thr Thr Arg Lys	
395	400	405
Gln Trp Leu Lys Glu Thr Pro Asp Thr	Leu Leu Asn Ile Leu Lys	
410	415	420
Asn Ala Asp Leu Ser Leu Ala Phe Gln	Thr Tyr Thr Ile Tyr Arg	
425	430	435
Pro Gly Ser Glu Gly Phe Leu Lys Gly	Pro Leu Ser Glu Glu Thr	
440	445	450
Glu Ala Ser Asp Ser Val Asp Gly Gly	His Asp Ser Val Ile Leu	
455	460	465
Asp Pro Glu Arg Leu Glu Pro Gly Leu	Asp Glu Glu Asp Thr Asp	
470	475	480
Phe Glu Glu Glu Asp Asp Asn Pro Asp	Trp Val Ser Glu Leu Lys	
485	490	495
Lys Arg Ala Gly Trp Gln Gly Leu Cys	Asp Asp Arg	
500	505	

&lt;210&gt; 8

&lt;211&gt; 2014

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7483337CD1

&lt;400&gt; 8

Met Glu Thr Leu Asn Gly Ala Gly Asp Thr Gly Gly Lys Pro Ser			
1	5	10	15
Thr Arg Gly Gly Asp Pro Ala Ala Arg Ser Arg Arg Thr Glu Gly			
20	25	30	
Ile Arg Ala Ala Tyr Arg Arg Gly Asp Arg Gly Gly Ala Arg Asp			
35	40	45	
Leu Leu Glu Glu Ala Cys Asp Gln Cys Ala Ser Gln Leu Glu Lys			
50	55	60	
Gly Gln Leu Leu Ser Ile Pro Ala Ala Tyr Gly Asp Leu Glu Met			
65	70	75	
Val Arg Tyr Leu Leu Ser Lys Arg Leu Val Glu Leu Pro Thr Glu			
80	85	90	
Pro Thr Asp Asp Asn Pro Ala Val Val Ala Ala Tyr Phe Gly His			
95	100	105	

Thr Ala Val Val Gln Asn Thr Leu Pro Thr Glu Pro Thr Asp Asp  
110 115 120  
Asn Pro Ala Val Val Ala Ala Tyr Phe Gly His Thr Ala Val Val  
125 130 135  
Gln Glu Leu Leu Glu Ser Leu Pro Gly Pro Cys Ser Pro Gln Arg  
140 145 150  
Leu Leu Asn Trp Met Leu Ala Leu Ala Cys Gln Arg Gly His Leu  
155 160 165  
Gly Val Val Lys Leu Leu Val Leu Thr His Gly Ala Asp Pro Glu  
170 175 180  
Ser Tyr Ala Val Arg Lys Asn Glu Phe Pro Val Ile Val Arg Leu  
185 190 195  
Pro Leu Tyr Ala Ala Ile Lys Ser Gly Asn Glu Asp Ile Ala Ile  
200 205 210  
Phe Leu Leu Arg His Gly Ala Tyr Phe Cys Ser Tyr Ile Leu Leu  
215 220 225  
Asp Ser Pro Asp Pro Ser Lys His Leu Leu Arg Lys Tyr Phe Ile  
230 235 240  
Glu Ala Ser Pro Leu Pro Ser Ser Tyr Pro Gly Lys Thr Ala Leu  
245 250 255  
Arg Val Lys Trp Ser His Leu Arg Leu Pro Trp Val Asp Leu Asp  
260 265 270  
Trp Leu Ile Asp Ile Ser Cys Gln Ile Thr Glu Leu Asp Leu Ser  
275 280 285  
Ala Asn Cys Leu Ala Thr Leu Pro Ser Val Ile Pro Trp Gly Leu  
290 295 300  
Ile Asn Leu Arg Lys Leu Asn Leu Ser Asp Asn His Leu Gly Glu  
305 310 315  
Leu Pro Gly Val Gln Ser Ser Asp Glu Ile Ile Cys Ser Arg Leu  
320 325 330  
Leu Glu Ile Asp Ile Ser Ser Asn Lys Leu Ser His Leu Pro Pro  
335 340 345  
Gly Phe Leu His Leu Ser Lys Leu Gln Lys Leu Thr Ala Ser Lys  
350 355 360  
Asn Cys Leu Glu Lys Leu Phe Glu Glu Asn Ala Thr Asn Trp  
365 370 375  
Ile Gly Leu Arg Lys Leu Gln Glu Leu Asp Ile Ser Asp Asn Lys  
380 385 390  
Leu Thr Glu Leu Pro Ala Leu Phe Leu His Ser Phe Lys Ser Leu  
395 400 405  
Asn Ser Leu Asn Val Ser Arg Asn Asn Leu Lys Val Phe Pro Asp  
410 415 420  
Pro Trp Ala Cys Pro Leu Lys Cys Cys Lys Ala Ser Arg Asn Ala  
425 430 435  
Leu Glu Cys Leu Pro Asp Lys Met Ala Val Phe Trp Lys Asn His  
440 445 450  
Leu Lys Asp Val Asp Phe Ser Glu Asn Ala Leu Lys Glu Val Pro  
455 460 465  
Leu Gly Leu Phe Gln Leu Asp Ala Leu Met Phe Leu Arg Leu Gln  
470 475 480  
Gly Asn Gln Leu Ala Ala Leu Pro Pro Gln Glu Lys Trp Thr Cys  
485 490 495  
Arg Gln Leu Lys Thr Leu Asp Leu Ser Arg Asn Gln Leu Gly Lys  
500 505 510  
Asn Glu Asp Gly Leu Lys Thr Lys Arg Ile Ala Phe Phe Thr Thr  
515 520 525

Arg Gly Arg Gln Arg Ser Gly Thr Glu Ala Glu Thr Thr Met Glu  
 530 535 540  
 Phe Ser Ala Ser Leu Val Thr Ile Val Phe Leu Ser Asn Asn Cys  
 545 550 555  
 Asn Leu Cys Ala Tyr Thr Cys Ala Ala Ser Val Leu Glu Phe Pro  
 560 565 570  
 Ala Phe Leu Ser Glu Ser Leu Glu Val Leu Cys Leu Asn Asp Asn  
 575 580 585  
 His Leu Asp Thr Val Pro Pro Ser Val Cys Leu Leu Lys Ser Leu  
 590 595 600  
 Ser Glu Leu Tyr Leu Gly Asn Asn Pro Gly Leu Arg Glu Leu Pro  
 605 610 615  
 Pro Glu Leu Gly Gln Leu Gly Asn Leu Trp Gln Leu Asp Thr Glu  
 620 625 630  
 Asp Leu Thr Ile Ser Asn Val Pro Ala Glu Ile Gln Lys Glu Gly  
 635 640 645  
 Pro Lys Ala Met Leu Ser Tyr Leu Arg Ala Gln Leu Arg Lys Ala  
 650 655 660  
 Glu Lys Cys Lys Leu Met Lys Met Ile Ile Val Gly Pro Pro Arg  
 665 670 675  
 Gln Gly Lys Ser Thr Leu Leu Glu Ile Leu Gln Thr Gly Arg Ala  
 680 685 690  
 Pro Gln Val Val His Gly Glu Ala Thr Ile Arg Thr Thr Lys Trp  
 695 700 705  
 Glu Leu Gln Arg Pro Ala Gly Ser Arg Ala Lys Val Lys Asp Gly  
 710 715 720  
 Leu Arg Ala Glu Ser Leu Trp Val Glu Ser Val Glu Phe Asn Val  
 725 730 735  
 Trp Asp Ile Gly Gly Pro Ala Ser Met Ala Thr Val Asn Gln Cys  
 740 745 750  
 Phe Phe Thr Asp Lys Ala Leu Tyr Val Val Val Trp Asn Leu Ala  
 755 760 765  
 Leu Gly Glu Ala Val Ala Asn Leu Gln Phe Trp Leu Leu Asn  
 770 775 780  
 Ile Glu Ala Lys Ala Pro Asn Ala Val Val Leu Val Val Gly Thr  
 785 790 795  
 His Leu Asp Leu Ile Glu Ala Lys Phe Arg Val Glu Arg Ile Ala  
 800 805 810  
 Thr Leu Arg Ala Tyr Val Leu Ala Leu Cys Arg Ser Pro Ser Gly  
 815 820 825  
 Ser Arg Ala Thr Gly Phe Pro Asp Ile Thr Phe Lys His Leu His  
 830 835 840  
 Glu Ile Ser Cys Lys Ser Leu Glu Gly Gln Glu Gly Leu Arg Gln  
 845 850 855  
 Leu Ile Phe His Val Thr Cys Ser Met Lys Asp Val Gly Ser Thr  
 860 865 870  
 Ile Gly Cys Gln Arg Leu Ala Gly Arg Leu Ile Pro Arg Ser Tyr  
 875 880 885  
 Leu Ser Leu Gln Glu Ala Val Leu Ala Glu Gln Gln Arg Arg Ser  
 890 895 900  
 Arg Asp Asp Asp Val Gln Tyr Leu Thr Asp Arg Gln Leu Glu Gln  
 905 910 915  
 Leu Val Glu Gln Thr Pro Asp Asn Asp Ile Lys Asp Tyr Glu Asp  
 920 925 930  
 Leu Gln Ser Ala Ile Ser Phe Leu Ile Glu Thr Gly Thr Leu Leu  
 935 940 945

His	Phe	Pro	Asp	Thr	Ser	His	Gly	Leu	Arg	Asn	Leu	Tyr	Phe	Leu
	950						955						960	
Asp	Pro	Ile	Trp	Leu	Ser	Glu	Cys	Leu	Gln	Arg	Ile	Phe	Asn	Ile
	965						970						975	
Lys	Gly	Ser	Arg	Ser	Val	Ala	Lys	Asn	Gly	Val	Ile	Arg	Ala	Glu
	980						985						990	
Asp	Leu	Arg	Met	Leu	Leu	Val	Gly	Thr	Gly	Phe	Thr	Gln	Gln	Thr
	995						1000						1005	
Glu	Glu	Gln	Tyr	Phe	Gln	Phe	Leu	Ala	Lys	Phe	Glu	Ile	Ala	Leu
	1010						1015						1020	
Pro	Val	Ala	Asn	Asp	Ser	Tyr	Leu	Leu	Pro	His	Leu	Leu	Pro	Ser
	1025						1030						1035	
Lys	Pro	Gly	Leu	Asp	Thr	His	Gly	Met	Arg	His	Pro	Thr	Ala	Asn
	1040						1045						1050	
Thr	Ile	Gln	Arg	Val	Phe	Lys	Met	Ser	Phe	Val	Pro	Val	Gly	Phe
	1055						1060						1065	
Trp	Gln	Arg	Phe	Ile	Ala	Arg	Met	Leu	Ile	Ser	Leu	Ala	Glu	Met
	1070						1075						1080	
Asp	Leu	Gln	Leu	Phe	Glu	Asn	Lys	Lys	Asn	Thr	Lys	Ser	Arg	Asn
	1085						1090						1095	
Arg	Lys	Val	Thr	Ile	Tyr	Ser	Phe	Thr	Gly	Asn	Gln	Arg	Asn	Arg
	1100						1105						1110	
Cys	Ser	Thr	Phe	Arg	Val	Lys	Arg	Asn	Gln	Thr	Ile	Tyr	Trp	Gln
	1115						1120						1125	
Glu	Gly	Leu	Leu	Val	Thr	Phe	Asp	Gly	Gly	Tyr	Leu	Ser	Val	Glu
	1130						1135						1140	
Ser	Ser	Asp	Val	Asn	Trp	Lys	Lys	Lys	Ser	Gly	Gly	Met	Lys	
	1145						1150						1155	
Ile	Val	Cys	Gln	Ser	Glu	Val	Arg	Asp	Phe	Ser	Ala	Met	Ala	Phe
	1160						1165						1170	
Ile	Thr	Asp	His	Val	Asn	Ser	Leu	Ile	Asp	Gln	Trp	Phe	Pro	Ala
	1175						1180						1185	
Leu	Thr	Ala	Thr	Glu	Ser	Asp	Gly	Thr	Pro	Leu	Met	Glu	Gln	Tyr
	1190						1195						1200	
Val	Pro	Cys	Pro	Val	Cys	Glu	Thr	Ala	Trp	Ala	Gln	His	Thr	Asp
	1205						1210						1215	
Pro	Ser	Glu	Lys	Ser	Glu	Asp	Val	Gln	Tyr	Phe	Asp	Met	Glu	Asp
	1220						1225						1230	
Cys	Val	Leu	Thr	Ala	Ile	Glu	Arg	Asp	Phe	Ile	Ser	Cys	Pro	Arg
	1235						1240						1245	
His	Pro	Asp	Leu	Pro	Val	Pro	Leu	Gln	Glu	Leu	Val	Pro	Glu	Leu
	1250						1255						1260	
Phe	Met	Thr	Asp	Phe	Pro	Ala	Arg	Leu	Phe	Leu	Glu	Asn	Ser	Lys
	1265						1270						1275	
Leu	Glu	His	Ser	Glu	Asp	Glu	Gly	Ser	Val	Leu	Gly	Gln	Gly	
	1280						1285						1290	
Ser	Gly	Thr	Val	Ile	Tyr	Arg	Ala	Arg	Tyr	Gln	Gly	Gln	Pro	Val
	1295						1300						1305	
Ala	Val	Lys	Arg	Phe	His	Ile	Lys	Lys	Phe	Lys	Asn	Phe	Ala	Asn
	1310						1315						1320	
Val	Pro	Ala	Asp	Thr	Met	Leu	Arg	His	Leu	Arg	Ala	Thr	Asp	Ala
	1325						1330						1335	
Met	Lys	Asn	Phe	Ser	Glu	Phe	Arg	Gln	Glu	Ala	Ser	Met	Leu	His
	1340						1345						1350	
Ala	Leu	Gln	His	Pro	Cys	Ile	Val	Ala	Leu	Ile	Gly	Ile	Ser	Ile
	1355						1360						1365	

His Pro Leu Cys Phe Ala Leu Glu Leu Ala Pro Leu Ser Ser Leu		
1370	1375	1380
Asn Thr Val Leu Ser Glu Asn Ala Arg Asp Ser Ser Phe Ile Pro		
1385	1390	1395
Leu Gly His Met Leu Thr Gln Lys Ile Ala Tyr Gln Ile Ala Ser		
1400	1405	1410
Gly Leu Ala Tyr Leu His Lys Lys Asn Ile Ile Phe Cys Asp Leu		
1415	1420	1425
Lys Ser Asp Asn Ile Leu Val Trp Ser Leu Asp Val Lys Glu His		
1430	1435	1440
Ile Asn Ile Lys Leu Ser Asp Tyr Gly Ile Ser Arg Gln Ser Phe		
1445	1450	1455
His Glu Gly Ala Leu Gly Val Glu Gly Thr Pro Gly Tyr Gln Ala		
1460	1465	1470
Pro Glu Ile Arg Pro Arg Ile Val Tyr Asp Glu Lys Val Asp Met		
1475	1480	1485
Phe Ser Tyr Gly Met Val Leu Tyr Glu Leu Leu Ser Gly Gln Arg		
1490	1495	1500
Pro Ala Leu Gly His His Gln Leu Gln Ile Ala Lys Lys Leu Ser		
1505	1510	1515
Lys Gly Ile Arg Pro Val Leu Gly Gln Pro Glu Glu Val Gln Phe		
1520	1525	1530
Arg Arg Leu Gln Ala Leu Met Met Glu Cys Trp Asp Thr Lys Pro		
1535	1540	1545
Glu Lys Arg Pro Leu Ala Leu Ser Val Val Ser Gln Met Lys Asp		
1550	1555	1560
Pro Thr Phe Ala Thr Phe Met Tyr Glu Leu Cys Cys Gly Lys Gln		
1565	1570	1575
Thr Ala Phe Phe Ser Ser Gln Gly Gln Glu Tyr Thr Val Val Phe		
1580	1585	1590
Trp Asp Gly Lys Glu Glu Ser Arg Asn Tyr Thr Val Val Asn Thr		
1595	1600	1605
Glu Lys Gly Leu Met Glu Val Gln Arg Met Cys Cys Pro Gly Met		
1610	1615	1620
Lys Val Ser Cys Gln Leu Gln Val Gln Arg Ser Leu Trp Thr Ala		
1625	1630	1635
Thr Glu Asn Ser Tyr Leu Val Leu Ala Gly Leu Ala Asp Gly Leu		
1640	1645	1650
Val Ala Val Phe Pro Val Val Arg Gly Thr Pro Lys Asp Ser Cys		
1655	1660	1665
Ser Tyr Leu Cys Ser His Thr Ala Asn Arg Ser Lys Phe Ser Ile		
1670	1675	1680
Ala Asp Glu Asp Ala Arg Gln Asn Pro Tyr Pro Val Lys Ala Met		
1685	1690	1695
Glu Val Val Asn Ser Gly Ser Glu Val Trp Tyr Ser Asn Gly Pro		
1700	1705	1710
Gly Leu Leu Val Ile Asp Cys Ala Ser Leu Glu Ile Cys Arg Arg		
1715	1720	1725
Leu Glu Pro Tyr Met Ala Pro Ser Met Val Thr Ser Val Val Cys		
1730	1735	1740
Ser Ser Glu Gly Arg Gly Glu Glu Val Val Trp Cys Leu Asp Asp		
1745	1750	1755
Lys Ala Asn Ser Leu Val Met Tyr His Ser Thr Thr Tyr Gln Leu		
1760	1765	1770
Cys Ala Arg Tyr Phe Cys Gly Val Pro Ser Pro Leu Arg Asp Met		
1775	1780	1785

Phe Pro Val Arg Pro Leu Asp Thr Glu Pro Pro Ala Ala Ser His  
 1790 1795 1800  
 Thr Ala Asn Pro Lys Val Pro Glu Gly Asp Ser Ile Ala Asp Val  
 1805 1810 1815  
 Ser Ile Met Tyr Ser Glu Glu Leu Gly Thr Gln Ile Leu Ile His  
 1820 1825 1830  
 Gln Glu Ser Leu Thr Asp Tyr Cys Ser Met Ser Ser Tyr Ser Ser  
 1835 1840 1845  
 Ser Pro Pro Arg Gln Ala Ala Arg Ser Pro Ser Ser Leu Pro Ser  
 1850 1855 1860  
 Ser Pro Ala Ser Ser Ser Val Pro Phe Ser Thr Asp Cys Glu  
 1865 1870 1875  
 Asp Ser Asp Met Leu His Thr Pro Gly Ala Ala Ser Asp Arg Ser  
 1880 1885 1890  
 Glu His Asp Leu Thr Pro Met Asp Gly Glu Thr Phe Ser Gln His  
 1895 1900 1905  
 Leu Gln Ala Val Lys Ile Leu Ala Val Arg Asp Leu Ile Trp Val  
 1910 1915 1920  
 Pro Arg Arg Gly Gly Asp Val Ile Val Ile Gly Leu Glu Lys Asp  
 1925 1930 1935  
 Ser Gly Ala Gln Arg Gly Arg Val Ile Ala Val Leu Lys Ala Arg  
 1940 1945 1950  
 Glu Leu Thr Pro His Gly Val Leu Val Asp Ala Ala Val Val Ala  
 1955 1960 1965  
 Lys Asp Thr Val Val Cys Thr Phe Glu Asn Glu Asn Thr Glu Trp  
 1970 1975 1980  
 Cys Leu Ala Val Trp Arg Gly Trp Gly Ala Arg Glu Phe Asp Ile  
 1985 1990 1995  
 Phe Tyr Gln Ser Tyr Glu Glu Leu Gly Arg Leu Glu Ala Cys Thr  
 2000 2005 2010  
 Arg Lys Arg Arg

<210> 9  
 <211> 348  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 6035509CD1

<400> 9  
 Met Met Leu Gly Leu Glu Ser Leu Pro Asp Pro Thr Asp Thr Trp  
 1 5 10 15  
 Glu Ile Ile Glu Thr Ile Gly Lys Gly Thr Tyr Gly Lys Val Tyr  
 20 25 30  
 Lys Val Thr Asn Lys Arg Asp Gly Ser Leu Ala Ala Val Lys Ile  
 35 40 45  
 Leu Asp Pro Val Ser Asp Met Asp Glu Glu Ile Glu Ala Glu Tyr  
 50 55 60  
 Asn Ile Leu Gln Phe Leu Pro Asn His Pro Asn Val Val Lys Phe  
 65 70 75  
 Tyr Gly Met Phe Tyr Lys Ala Asp His Cys Val Gly Gly Gln Leu  
 80 85 90  
 Trp Leu Val Leu Glu Leu Cys Asn Gly Gly Ser Val Thr Glu Leu

95	100	105
Val Lys Gly Leu Leu Arg Cys Gly Gln	Arg Leu Asp Glu Ala	Met
110	115	120
Ile Ser Tyr Ile Leu Tyr Gly Ala Leu	Leu Gly Leu Gln His	Leu
125	130	135
His Asn Asn Arg Ile Ile His Arg Asp	Val Lys Gly Asn Asn	Ile
140	145	150
Leu Leu Thr Thr Glu Gly Gly Val Lys	Leu Val Asp Phe Gly	Val
155	160	165
Ser Ala Gln Leu Thr Ser Thr Arg Leu	Arg Arg Asn Thr Ser	Val
170	175	180
Gly Thr Pro Phe Trp Met Ala Pro Glu	Val Ile Ala Cys Glu	Gln
185	190	195
Gln Tyr Asp Ser Ser Tyr Asp Ala Arg	Cys Asp Val Trp Ser	Leu
200	205	210
Gly Ile Thr Ala Ile Glu Leu Gly Asp	Gly Asp Pro Pro	Leu Phe
215	220	225
Asp Met His Pro Val Lys Thr Leu Phe	Lys Ile Pro Arg Asn	Pro
230	235	240
Pro Pro Thr Leu Leu His Pro Glu Lys	Trp Cys Glu Glu Phe	Asn
245	250	255
His Phe Ile Ser Gln Cys Leu Ile Lys	Asp Phe Glu Arg Arg	Pro
260	265	270
Ser Val Thr His Leu Leu Asp His Pro	Phe Ile Lys Gly Val	His
275	280	285
Gly Lys Val Leu Phe Leu Gln Lys Gln	Leu Ala Lys Val Leu	Gln
290	295	300
Asp Gln Lys His Gln Asn Pro Val Ala	Lys Thr Arg His Glu	Arg
305	310	315
Met His Thr Arg Arg Pro Tyr His Val	Glu Asp Ala Glu Lys	Tyr
320	325	330
Cys Leu Glu Asp Asp Leu Val Asn Leu	Glu Val Leu Asp Glu	Val
335	340	345
Leu Asn Ile		

<210> 10  
 <211> 2042  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7373485CD1

<400> 10

Met Ala Thr Asp Asp Lys Thr Ser Pro	Thr Leu Asp Ser Ala Asn		
1	5	10	15
Asp Leu Pro Arg Ser Pro	Thr Ser Pro Ser His	Leu Thr His Phe	
20	25	30	
Lys Pro Leu Thr Pro Asp Gln Asp	Glu Pro Pro Phe	Lys Ser Ala	
35	40	45	
Tyr Ser Ser Phe Val Asn Leu Phe Arg	Phe Asn Lys Glu Arg	Ala	
50	55	60	
Glu Gly Gln Gly Glu Gln Gln Pro	Leu Ser Gly Ser Trp	Thr	
65	70	75	

Ser Pro Gln Leu Pro Ser Arg Thr Gln Ser Val Arg Ser Pro Thr  
80 85 90  
Pro Tyr Lys Lys Gln Leu Asn Glu Glu Leu Gln Arg Arg Ser Ser  
95 100 105  
Ala Leu Asp Thr Arg Arg Lys Ala Glu Pro Thr Phe Gly Gly His  
110 115 120  
Asp Pro Arg Thr Ala Val Gln Leu Arg Ser Leu Ser Thr Val Leu  
125 130 135  
Lys Arg Leu Lys Glu Ile Met Glu Gly Lys Ser Gln Asp Ser Asp  
140 145 150  
Leu Lys Gln Tyr Trp Met Pro Asp Ser Gln Cys Lys Glu Cys Tyr  
155 160 165  
Asp Cys Ser Glu Lys Phe Thr Thr Phe Arg Arg His His Cys  
170 175 180  
Arg Leu Cys Gly Gln Ile Phe Cys Ser Arg Cys Cys Asn Gln Glu  
185 190 195  
Ile Pro Gly Lys Phe Met Gly Tyr Thr Gly Asp Leu Arg Ala Cys  
200 205 210  
Thr Tyr Cys Arg Lys Ile Ala Leu Ser Tyr Ala His Ser Thr Asp  
215 220 225  
Ser Asn Ser Ile Gly Glu Asp Leu Asn Ala Leu Ser Asp Ser Ala  
230 235 240  
Cys Ser Val Ser Val Leu Asp Pro Ser Glu Pro Arg Thr Pro Val  
245 250 255  
Gly Ser Arg Lys Ala Ser Arg Asn Ile Phe Leu Glu Asp Asp Leu  
260 265 270  
Ala Trp Gln Ser Leu Ile His Pro Asp Ser Ser Asn Thr Pro Leu  
275 280 285  
Ser Thr Arg Leu Val Ser Val Gln Glu Asp Ala Gly Lys Ser Pro  
290 295 300  
Ala Arg Asn Arg Ser Ala Ser Ile Thr Asn Leu Ser Leu Asp Arg  
305 310 315  
Ser Gly Ser Pro Met Val Pro Ser Tyr Glu Thr Ser Val Ser Pro  
320 325 330  
Gln Ala Asn Arg Thr Tyr Val Arg Thr Glu Thr Thr Glu Asp Glu  
335 340 345  
Arg Lys Ile Leu Leu Asp Ser Val Gln Leu Lys Asp Leu Trp Lys  
350 355 360  
Lys Ile Cys His His Ser Ser Gly Met Glu Phe Gln Asp His Arg  
365 370 375  
Tyr Trp Leu Arg Thr His Pro Asn Cys Ile Val Gly Lys Glu Leu  
380 385 390  
Val Asn Trp Leu Ile Arg Asn Gly His Ile Ala Thr Arg Ala Gln  
395 400 405  
Ala Ile Ala Ile Gly Gln Ala Met Val Asp Gly Arg Trp Leu Asp  
410 415 420  
Cys Val Ser His His Asp Gln Leu Phe Arg Asp Glu Tyr Ala Leu  
425 430 435  
Tyr Arg Pro Leu Gln Ser Thr Glu Phe Ser Glu Thr Pro Ser Pro  
440 445 450  
Asp Ser Asp Ser Val Asn Ser Val Glu Gly His Ser Glu Pro Ser  
455 460 465  
Trp Phe Lys Asp Ile Lys Phe Asp Asp Ser Asp Thr Glu Gln Ile  
470 475 480  
Ala Glu Glu Gly Asp Asp Asn Leu Ala Lys Tyr Leu Ile Ser Asp  
485 490 495

Thr Gly Gly Gln Gln Leu Ser Ile Ser Asp Ala Phe Ile Lys Glu  
 500 505 510  
 Ser Leu Phe Asn Arg Arg Val Glu Glu Lys Ser Lys Glu Leu Pro  
 515 520 525  
 Phe Thr Pro Leu Gly Trp His His Asn Asn Leu Glu Leu Leu Arg  
 530 535 540  
 Glu Glu Asn Gly Glu Lys Gln Ala Met Glu Arg Leu Leu Ser Ala  
 545 550 555  
 Asn His Asn His Met Met Ala Leu Leu Gln Gln Leu Leu His Ser  
 560 565 570  
 Asp Ser Leu Ser Ser Ser Trp Arg Asp Ile Ile Val Ser Leu Val  
 575 580 585  
 Cys Gln Val Val Gln Thr Val Arg Pro Asp Val Lys Asn Gln Asp  
 590 595 600  
 Asp Asp Met Asp Ile Arg Gln Phe Val His Ile Lys Lys Ile Pro  
 605 610 615  
 Gly Gly Lys Lys Phe Asp Ser Val Val Val Asn Gly Phe Val Cys  
 620 625 630  
 Thr Lys Asn Ile Ala His Lys Lys Met Asn Ser Cys Ile Lys Asn  
 635 640 645  
 Pro Lys Ile Leu Leu Leu Lys Cys Ser Ile Glu Tyr Leu Tyr Arg  
 650 655 660  
 Glu Glu Thr Lys Phe Thr Cys Ile Asp Pro Ile Val Leu Gln Glu  
 665 670 675  
 Arg Glu Phe Leu Lys Asn Tyr Val Gln Arg Ile Val Asp Val Arg  
 680 685 690  
 Pro Thr Leu Val Leu Val Glu Lys Thr Val Ser Arg Ile Ala Gln  
 695 700 705  
 Asp Met Leu Leu Glu His Gly Ile Thr Leu Val Ile Asn Val Lys  
 710 715 720  
 Ser Gln Val Leu Glu Arg Ile Ser Arg Met Thr Gln Gly Asp Leu  
 725 730 735  
 Val Met Ser Met Asp Gln Leu Leu Thr Lys Pro Arg Leu Gly Thr  
 740 745 750  
 Cys His Lys Phe Tyr Met Gln Ile Phe Gln Leu Pro Asn Glu Gln  
 755 760 765  
 Thr Lys Thr Leu Met Phe Phe Glu Gly Cys Pro Gln His Leu Gly  
 770 775 780  
 Cys Thr Ile Lys Leu Arg Gly Gly Ser Asp Tyr Glu Leu Ala Arg  
 785 790 795  
 Val Lys Glu Ile Leu Ile Phe Met Ile Cys Val Ala Tyr His Ser  
 800 805 810  
 Gln Leu Glu Ile Ser Phe Leu Met Asp Glu Phe Ala Met Pro Pro  
 815 820 825  
 Thr Leu Met Gln Asn Pro Ser Phe His Ser Leu Ile Glu Gly Arg  
 830 835 840  
 Gly His Glu Gly Ala Val Gln Glu Gln Tyr Gly Gly Ser Ile  
 845 850 855  
 Pro Trp Asp Pro Asp Ile Pro Pro Glu Ser Leu Pro Cys Asp Asp  
 860 865 870  
 Ser Ser Leu Leu Glu Ser Arg Ile Val Phe Glu Lys Gly Glu Gln  
 875 880 885  
 Glu Asn Lys Asn Leu Pro Gln Ala Val Ala Ser Val Lys His Gln  
 890 895 900  
 Glu His Ser Thr Thr Ala Cys Pro Ala Gly Leu Pro Cys Ala Phe  
 905 910 915

Phe Ala Pro Val Pro Glu Ser Leu Leu Pro Leu Pro Val Asp Asp  
 920 925 930  
 Gln Gln Asp Ala Leu Gly Ser Glu Leu Pro Glu Ser Leu Gln Gln  
 935 940 945  
 Thr Val Val Leu Gln Asp Pro Lys Ser Gln Ile Arg Ala Phe Arg  
 950 955 960  
 Asp Pro Leu Gln Asp Asp Thr Gly Leu Tyr Val Thr Glu Glu Val  
 965 970 975  
 Thr Ser Ser Glu Asp Lys Arg Lys Thr Tyr Ser Leu Ala Phe Lys  
 980 985 990  
 Gln Glu Leu Lys Asp Val Ile Leu Cys Ile Ser Pro Val Ile Thr  
 995 1000 1005  
 Phe Arg Glu Pro Phe Leu Leu Thr Glu Lys Gly Met Arg Cys Ser  
 1010 1015 1020  
 Thr Arg Asp Tyr Phe Ala Glu Gln Val Tyr Trp Ser Pro Leu Leu  
 1025 1030 1035  
 Asn Lys Glu Phe Lys Glu Met Glu Asn Arg Arg Lys Lys Gln Leu  
 1040 1045 1050  
 Leu Arg Asp Leu Ser Gly Leu Gln Gly Met Asn Gly Ser Ile Gln  
 1055 1060 1065  
 Ala Lys Ser Ile Gln Val Leu Pro Ser His Glu Leu Val Ser Thr  
 1070 1075 1080  
 Arg Ile Ala Glu His Leu Gly Asp Ser Gln Ser Leu Gly Arg Met  
 1085 1090 1095  
 Leu Ala Asp Tyr Arg Ala Arg Gly Gly Arg Ile Gln Pro Lys Asn  
 1100 1105 1110  
 Ser Asp Pro Phe Ala His Ser Lys Asp Ala Ser Ser Thr Ser Ser  
 1115 1120 1125  
 Gly Lys Ser Gly Ser Lys Asn Glu Gly Asp Glu Glu Arg Gly Leu  
 1130 1135 1140  
 Ile Leu Ser Asp Ala Val Trp Ser Thr Lys Val Asp Cys Leu Asn  
 1145 1150 1155  
 Pro Ile Asn His Gln Arg Leu Cys Val Leu Phe Ser Ser Ser Ser  
 1160 1165 1170  
 Ala Gln Ser Ser Asn Ala Pro Ser Ala Cys Val Ser Pro Trp Ile  
 1175 1180 1185  
 Val Thr Met Glu Phe Tyr Gly Lys Asn Asp Leu Thr Leu Gly Ile  
 1190 1195 1200  
 Phe Leu Glu Arg Tyr Cys Phe Arg Pro Ser Tyr Gln Cys Pro Ser  
 1205 1210 1215  
 Met Phe Cys Asp Thr Pro Met Val His His Ile Arg Arg Phe Val  
 1220 1225 1230  
 His Gly Gln Gly Cys Val Gln Ile Ile Leu Lys Glu Leu Asp Ser  
 1235 1240 1245  
 Pro Val Pro Gly Tyr Gln His Thr Ile Leu Thr Tyr Ser Trp Cys  
 1250 1255 1260  
 Arg Ile Cys Lys Gln Val Thr Pro Val Val Ala Leu Ser Asn Glu  
 1265 1270 1275  
 Ser Trp Ser Met Ser Phe Ala Lys Tyr Leu Glu Leu Arg Phe Tyr  
 1280 1285 1290  
 Gly His Gln Tyr Thr Arg Arg Ala Asn Ala Glu Pro Cys Gly His  
 1295 1300 1305  
 Ser Ile His His Asp Tyr His Gln Tyr Phe Ser Tyr Asn Gln Met  
 1310 1315 1320  
 Val Ala Ser Phe Ser Tyr Ser Pro Ile Arg Leu Leu Glu Val Cys  
 1325 1330 1335

Val Pro Leu Pro Lys Ile Phe Ile Lys Arg Gln Ala Pro Leu Lys  
 1340 1345 1350  
 Val Ser Leu Leu Gln Asp Leu Lys Asp Phe Phe Gln Lys Val Ser  
 1355 1360 1365  
 Gln Val Tyr Val Ala Ile Asp Glu Arg Leu Ala Ser Leu Lys Thr  
 1370 1375 1380  
 Asp Thr Phe Ser Lys Thr Arg Glu Glu Lys Met Glu Asp Ile Phe  
 1385 1390 1395  
 Ala Gln Lys Glu Met Glu Glu Gly Glu Phe Lys Asn Trp Ile Glu  
 1400 1405 1410  
 Lys Met Gln Ala Arg Leu Met Ser Ser Val Asp Thr Pro Gln  
 1415 1420 1425  
 Gln Leu Gln Ser Val Phe Glu Ser Leu Ile Ala Lys Lys Gln Ser  
 1430 1435 1440  
 Leu Cys Glu Val Leu Gln Ala Trp Asn Asn Arg Leu Gln Asp Leu  
 1445 1450 1455  
 Phe Gln Gln Glu Lys Gly Arg Lys Arg Pro Ser Val Pro Pro Ser  
 1460 1465 1470  
 Pro Gly Arg Leu Arg Gln Gly Glu Glu Ser Lys Ile Ser Ala Met  
 1475 1480 1485  
 Asp Ala Ser Pro Arg Asn Ile Ser Pro Gly Leu Gln Asn Gly Glu  
 1490 1495 1500  
 Lys Glu Asp Arg Phe Leu Thr Thr Leu Ser Ser Gln Ser Ser Thr  
 1505 1510 1515  
 Ser Ser Thr His Leu Gln Leu Pro Thr Pro Pro Glu Val Met Ser  
 1520 1525 1530  
 Glu Gln Ser Val Gly Gly Pro Pro Glu Leu Asp Thr Ala Ser Ser  
 1535 1540 1545  
 Ser Glu Asp Val Phe Asp Gly His Leu Leu Gly Ser Thr Asp Ser  
 1550 1555 1560  
 Gln Val Lys Glu Lys Ser Thr Met Lys Ala Ile Phe Ala Asn Leu  
 1565 1570 1575  
 Leu Pro Gly Asn Ser Tyr Asn Pro Ile Pro Phe Pro Phe Asp Pro  
 1580 1585 1590  
 Asp Lys His Tyr Leu Met Tyr Glu His Glu Arg Val Pro Ile Ala  
 1595 1600 1605  
 Val Cys Glu Lys Glu Pro Ser Ser Ile Ile Ala Phe Ala Leu Ser  
 1610 1615 1620  
 Cys Lys Glu Tyr Arg Asn Ala Leu Glu Glu Leu Ser Lys Ala Thr  
 1625 1630 1635  
 Gln Trp Asn Ser Ala Glu Glu Gly Leu Pro Thr Asn Ser Thr Ser  
 1640 1645 1650  
 Asp Ser Arg Pro Lys Ser Ser Ser Pro Ile Arg Leu Pro Glu Met  
 1655 1660 1665  
 Ser Gly Gly Gln Thr Asn Arg Thr Thr Glu Thr Glu Pro Gln Pro  
 1670 1675 1680  
 Thr Lys Lys Ala Ser Gly Met Leu Ser Phe Phe Arg Gly Thr Ala  
 1685 1690 1695  
 Gly Lys Ser Pro Asp Leu Ser Ser Gln Lys Arg Glu Thr Leu Arg  
 1700 1705 1710  
 Gly Ala Asp Ser Ala Tyr Tyr Gln Val Gly Gln Thr Gly Lys Glu  
 1715 1720 1725  
 Gly Thr Glu Asn Gln Gly Val Glu Pro Gln Asp Glu Val Asp Gly  
 1730 1735 1740  
 Gly Asp Thr Gln Lys Lys Gln Leu Ile Asn Pro His Val Glu Leu  
 1745 1750 1755

Gln Phe Ser Asp Ala Asn Ala Lys Phe Tyr Cys Arg Leu Tyr Tyr  
 1760 1765 1770  
 Ala Gly Glu Phe His Lys Met Arg Glu Val Ile Leu Asp Ser Ser  
 1775 1780 1785  
 Glu Glu Asp Phe Ile Arg Ser Leu Ser His Ser Ser Pro Trp Gln  
 1790 1795 1800  
 Ala Arg Gly Gly Lys Ser Gly Ala Ala Phe Tyr Ala Thr Glu Asp  
 1805 1810 1815  
 Asp Arg Phe Ile Leu Lys Gln Met Pro Arg Leu Glu Val Gln Ser  
 1820 1825 1830  
 Phe Leu Asp Phe Ala Pro His Tyr Phe Asn Tyr Ile Thr Asn Ala  
 1835 1840 1845  
 Val Gln Gln Lys Arg Pro Thr Ala Leu Ala Lys Ile Leu Gly Val  
 1850 1855 1860  
 Tyr Arg Ile Gly Tyr Lys Asn Ser Gln Asn Asn Thr Glu Lys Lys  
 1865 1870 1875  
 Leu Asp Leu Leu Val Met Glu Asn Leu Phe Tyr Gly Arg Lys Met  
 1880 1885 1890  
 Ala Gln Val Phe Asp Leu Lys Gly Ser Leu Arg Asn Arg Asn Val  
 1895 1900 1905  
 Lys Thr Asp Thr Gly Lys Glu Ser Cys Asp Val Val Leu Leu Asp  
 1910 1915 1920  
 Glu Asn Leu Leu Lys Met Val Arg Asp Asn Pro Leu Tyr Ile Arg  
 1925 1930 1935  
 Ser His Ser Lys Ala Val Leu Arg Thr Ser Ile His Ser Asp Ser  
 1940 1945 1950  
 His Phe Leu Ser Ser His Leu Ile Ile Asp Tyr Ser Leu Leu Val  
 1955 1960 1965  
 Gly Arg Asp Asp Thr Ser Asn Glu Leu Val Val Gly Ile Ile Asp  
 1970 1975 1980  
 Tyr Ile Arg Thr Phe Thr Trp Asp Lys Leu Glu Met Val Val  
 1985 1990 1995  
 Lys Ser Thr Gly Ile Leu Gly Gly Gln Gly Lys Met Pro Thr Val  
 2000 2005 2010  
 Val Ser Pro Glu Leu Tyr Arg Thr Arg Phe Cys Glu Ala Met Asp  
 2015 2020 2025  
 Lys Tyr Phe Leu Met Val Pro Asp His Trp Thr Gly Leu Gly Leu  
 2030 2035 2040

Asn Cys

<210> 11  
 <211> 551  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5734965CD1

<400> 11  
 Met Ser Gly Gly Glu Gln Lys Pro Glu Arg Tyr Tyr Val Gly Val  
 1 5 10 15  
 Asp Val Gly Thr Gly Ser Val Arg Ala Ala Leu Val Asp Gln Ser  
 20 25 30  
 Gly Val Leu Leu Ala Phe Ala Asp Gln Pro Ile Lys Asn Trp Glu

35	40	45
Pro Gln Phe Asn His His Glu Gln Ser Ser	Glu Asp Ile Trp Ala	
50	55	60
Ala Cys Cys Val Val Thr Lys Lys Val Val	Gln Gly Ile Asp Leu	
65	70	75
Asn Gln Ile Arg Gly Leu Gly Phe Asp Ala	Thr Cys Ser Leu Val	
80	85	90
Val Leu Asp Lys Gln Phe His Pro Leu Pro	Val Asn Gln Glu Gly	
95	100	105
Asp Ser His Arg Asn Val Ile Met Trp Leu	Asp His Arg Ala Val	
110	115	120
Ser Gln Val Asn Arg Ile Asn Glu Thr Lys	His Ser Val Leu Gln	
125	130	135
Tyr Val Gly Gly Val Met Ser Val Glu Met	Gln Ala Pro Lys Leu	
140	145	150
Leu Trp Leu Lys Glu Asn Leu Arg Glu Ile	Cys Trp Asp Lys Ala	
155	160	165
Gly His Phe Phe Asp Leu Pro Asp Phe Leu	Ser Trp Lys Ala Thr	
170	175	180
Gly Val Thr Ala Arg Ser Leu Cys Ser Leu	Val Cys Lys Trp Thr	
185	190	195
Tyr Ser Ala Glu Lys Gly Trp Asp Asp Ser	Phe Trp Lys Met Ile	
200	205	210
Gly Leu Glu Asp Phe Val Ala Asp Asn Tyr	Ser Lys Ile Gly Asn	
215	220	225
Gln Val Leu Pro Pro Gly Ala Ser Leu Gly	Asn Gln Gly Leu Thr Pro	
230	235	240
Glu Ala Ala Arg Asp Leu Gly Leu Leu Pro	Gly Ile Ala Val Ala	
245	250	255
Ala Ser Leu Ile Asp Ala His Ala Gly Gly	Leu Gly Val Ile Gly	
260	265	270
Ala Asp Val Arg Gly His Gly Leu Ile Cys	Glu Gly Gln Pro Val	
275	280	285
Thr Ser Arg Leu Ala Val Ile Cys Gly Thr	Ser Ser Cys His Met	
290	295	300
Gly Ile Ser Lys Asp Pro Ile Phe Val Pro	Gly Val Trp Gly Pro	
305	310	315
Tyr Phe Ser Ala Met Val Pro Gly Phe Trp	Leu Asn Glu Gly Gly	
320	325	330
Gln Ser Val Thr Gly Lys Leu Ile Asp His	Met Val Gln Gly His	
335	340	345
Ala Ala Phe Pro Glu Leu Gln Val Lys Ala	Thr Ala Arg Cys Gln	
350	355	360
Ser Ile Tyr Ala Tyr Leu Asn Ser His Leu	Asp Leu Ile Lys Lys	
365	370	375
Ala Gln Pro Val Gly Phe Leu Thr Val Asp	Leu His Val Trp Pro	
380	385	390
Asp Phe His Gly Asn Arg Ser Pro Leu Ala	Asp Leu Thr Leu Lys	
395	400	405
Gly Met Val Thr Gly Leu Lys Leu Ser Gln	Asp Leu Asp Asp Leu	
410	415	420
Ala Ile Leu Tyr Leu Ala Thr Val Gln Ala	Ile Ala Leu Gly Thr	
425	430	435
Arg Phe Ile Ile Glu Ala Met Glu Ala Ala	Gly His Ser Ile Ser	
440	445	450
Thr Leu Phe Leu Cys Gly Leu Ser Lys Asn	Pro Leu Phe Val	

455	460	465
Gln Met His Ala Asp Ile Thr Gly Met	Pro Val Val Leu Ser	Gln
470	475	480
Glu Val Glu Ser Val Leu Val Gly Ala	Ala Val Leu Gly Ala	Cys
485	490	495
Ala Ser Gly Asp Phe Ala Ser Val Gln	Glu Ala Met Ala Lys	Met
500	505	510
Ser Lys Val Gly Lys Val Val Phe Pro	Arg Leu Gln Asp Lys	Lys
515	520	525
Tyr Tyr Asp Lys Lys Tyr Gln Val Phe	Leu Lys Leu Val Glu	His
530	535	540
Gln Lys Glu Tyr Leu Ala Ile Met Asn Asp Asp		
545	550	

&lt;210&gt; 12

&lt;211&gt; 485

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473788CD1

&lt;400&gt; 12

Met Arg Ser Gly Ala Glu Arg Arg Gly Ser Ser Ala Ala Ala Ser		
1	5	10
Pro Gly Ser Pro Pro Pro Gly Arg Ala Arg Pro Ala Gly Ser Asp		
20	25	30
Ala Pro Ser Ala Leu Pro Pro Pro Ala Ala Gly Gln Pro Arg Ala		
35	40	45
Arg Asp Ser Gly Asp Val Arg Ser Gln Pro Arg Pro Leu Phe Gln		
50	55	60
Trp Ser Lys Trp Lys Lys Arg Met Gly Ser Ser Met Ser Ala Ala		
65	70	75
Thr Ala Arg Arg Pro Val Phe Asp Asp Lys Glu Asp Val Asn Phe		
80	85	90
Asp His Phe Gln Ile Leu Arg Ala Ile Gly Lys Gly Ser Phe Gly		
95	100	105
Lys Val Cys Ile Val Gln Lys Arg Asp Thr Glu Lys Met Tyr Ala		
110	115	120
Met Lys Tyr Met Asn Lys Gln Gln Cys Ile Glu Arg Asp Glu Val		
125	130	135
Arg Asn Val Phe Arg Glu Leu Glu Ile Leu Gln Glu Ile Glu His		
140	145	150
Val Phe Leu Val Asn Leu Trp Tyr Ser Phe Gln Asp Glu Glu Asp		
155	160	165
Met Phe Met Val Val Asp Leu Leu Leu Gly Gly Asp Leu Arg Tyr		
170	175	180
His Leu Gln Gln Asn Val Gln Phe Ser Glu Asp Thr Val Arg Leu		
185	190	195
Tyr Ile Cys Glu Met Ala Leu Ala Leu Asp Tyr Leu Arg Gly Gln		
200	205	210
His Ile Ile His Arg Asp Val Lys Pro Asp Asn Ile Leu Leu Asp		
215	220	225
Glu Arg Gly His Ala His Leu Thr Asp Phe Asn Ile Ala Thr Ile		
230	235	240

Ile Lys Asp Gly Glu Arg Ala Thr Ala Leu Ala Gly Thr Lys Pro  
 245 250 255  
 Tyr Met Ala Pro Glu Ile Phe His Ser Phe Val Asn Gly Gly Thr  
 260 265 270  
 Gly Tyr Ser Phe Glu Val Asp Trp Trp Ser Val Gly Val Met Ala  
 275 280 285  
 Tyr Glu Leu Leu Arg Gly Trp Arg Pro Tyr Asp Ile His Ser Ser  
 290 295 300  
 Asn Ala Val Glu Ser Leu Val Gln Leu Phe Ser Thr Val Ser Val  
 305 310 315  
 Gln Tyr Val Pro Thr Trp Ser Lys Glu Met Val Ala Leu Leu Arg  
 320 325 330  
 Lys Leu Leu Thr Val Asn Pro Glu His Arg Leu Ser Ser Leu Gln  
 335 340 345  
 Asp Val Gln Ala Ala Pro Ala Leu Ala Gly Val Leu Trp Asp His  
 350 355 360  
 Leu Ser Glu Lys Arg Val Glu Pro Gly Phe Val Pro Asn Lys Gly  
 365 370 375  
 Arg Leu His Cys Asp Pro Thr Phe Glu Leu Glu Glu Met Ile Leu  
 380 385 390  
 Glu Ser Arg Pro Leu His Lys Lys Lys Arg Leu Ala Lys Asn  
 395 400 405  
 Lys Ser Arg Asp Asn Ser Arg Asp Ser Ser Gln Ser Glu Asn Asp  
 410 415 420  
 Tyr Leu Gln Asp Cys Leu Asp Ala Ile Gln Gln Asp Phe Val Ile  
 425 430 435  
 Phe Asn Arg Glu Lys Leu Lys Arg Ser Gln Asp Leu Pro Arg Glu  
 440 445 450  
 Pro Leu Pro Ala Leu Ser Pro Gly Met Leu Arg Ser Leu Trp Arg  
 455 460 465  
 Thr Arg Arg Thr Leu Arg Leu Pro Met Cys Gly Pro Ile Cys Pro  
 470 475 480  
 Ser Ala Gly Ser Gly  
 485

<210> 13

<211> 282

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3107989CD1

<400> 13

Met Pro Ala Phe Ile Gln Met Gly Arg Asp Lys Asn Phe Ser Ser  
 1 5 10 15  
 Leu His Thr Val Phe Cys Ala Thr Gly Gly Ala Tyr Lys Phe  
 20 25 30  
 Glu Gln Asp Phe Leu Thr Ile Gly Asp Leu Gln Leu Cys Lys Leu  
 35 40 45  
 Asp Glu Leu Asp Cys Leu Ile Lys Gly Ile Leu Tyr Ile Asp Ser  
 50 55 60  
 Val Gly Phe Asn Gly Arg Ser Gln Cys Tyr Tyr Phe Glu Asn Pro  
 65 70 75  
 Ala Asp Ser Glu Lys Cys Gln Lys Leu Pro Phe Asp Leu Lys Asn

80	85	90
Pro Tyr Pro Leu Leu Leu Val Asn Ile	Gly Ser Gly Val Ser	Ile
95	100	105
Leu Ala Val Tyr Ser Lys Asp Asn Tyr	Lys Arg Val Thr Gly	Thr
110	115	120
Ser Leu Gly Gly Thr Phe Phe Gly	Leu Cys Cys Leu Leu	Thr
125	130	135
Gly Cys Thr Thr Phe Glu Glu Ala Leu	Glu Met Ala Ser Arg	Gly
140	145	150
Asp Ser Thr Lys Val Asp Lys Leu Val	Arg Asp Ile Tyr Gly	Gly
155	160	165
Asp Tyr Glu Arg Phe Gly Leu Pro Gly	Trp Ala Val Ala Ser	Ser
170	175	180
Phe Gly Asn Met Met Ser Lys Glu Lys	Arg Asp Ser Ile Ser	Lys
185	190	195
Glu Asp Leu Ala Arg Ala Thr Leu Val	Thr Ile Thr Asn Asn	Ile
200	205	210
Gly Ser Ile Ala Arg Met Cys Ala Leu	Asn Glu Asn Ile Asp	Arg
215	220	225
Val Val Phe Val Gly Asn Phe Leu Arg	Ile Asn Met Val Ser	Met
230	235	240
Lys Leu Leu Ala Tyr Ala Met Asp Phe	Trp Ser Lys Gly Gln	Leu
245	250	255
Lys Ala Leu Phe Leu Glu His Glu Gly	Tyr Phe Gly Ala Val	Gly
260	265	270
Ala Leu Leu Glu Leu Phe Lys Met Thr	Asp Asp Lys	
275	280	

<210> 14  
 <211> 151  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7482887CD1

<400> 14

Met Ala Asn Thr Glu Ser Ile Ile Ile	Asn Pro Ser Ala Val Gln		
1	5	10	15
His Ser Leu Val Gly Glu Ile Ile Lys	Tyr Ser Glu Gln Lys Gly		
20	25	30	
Phe Tyr Leu Val Thr Met Lys Phe Leu Arg	Ala Ser Glu Lys Pro		
35	40	45	
Leu Lys Pro His Tyr Thr Asn Leu Lys Asp	His Pro Phe Phe Pro		
50	55	60	
Asp Leu Val Lys Tyr Met Asn Ser Gly Gln	Val Val Ala Met Val		
65	70	75	
Leu Glu Gly Leu Asn Val Ala Lys Thr Gly	Leu Arg Met Leu Gly		
80	85	90	
Glu Thr Asn Ser Leu Gly Ser Met Leu	Glu Thr Ile Ile Arg Arg		
95	100	105	
Asp Phe Cys Ala Lys Ile Gly Gly Asn Val	Ile Gly Gly Ser Asp		
110	115	120	
Ser Leu Gln Ser Ala Glu Lys Glu Ile	Ser Leu Trp Phe Lys Pro		
125	130	135	

Lys Glu Pro Val Asp Tyr Arg Ser Cys Ala Tyr Asp Trp Val Tyr  
140 145 150  
Ala

<210> 15  
<211> 410  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2963414CD1

<400> 15  
Met Val Val Gln Asn Ser Ala Asp Ala Gly Asp Met Arg Ala Gly  
1 5 10 15  
Val Gln Leu Glu Pro Phe Leu His Gln Val Gly Gly His Met Ser  
20 25 30  
Val Met Lys Tyr Asp Glu His Thr Val Cys Lys Pro Leu Val Ser  
35 40 45  
Arg Glu Gln Arg Phe Tyr Glu Ser Leu Pro Leu Ala Met Lys Arg  
50 55 60  
Phe Thr Pro Gln Tyr Lys Gly Thr Val Thr Val His Leu Trp Lys  
65 70 75  
Asp Ser Thr Gly His Leu Ser Leu Val Ala Asn Pro Val Lys Glu  
80 85 90  
Ser Gln Glu Pro Phe Lys Val Ser Thr Glu Ser Ala Ala Val Ala  
95 100 105  
Ile Trp Gln Thr Leu Gln Gln Thr Thr Gly Ser Asn Gly Ser Asp  
110 115 120  
Cys Thr Leu Ala Gln Trp Pro His Ala Gln Leu Ala Arg Ser Pro  
125 130 135  
Lys Glu Ser Pro Ala Lys Ala Leu Leu Arg Ser Glu Pro His Leu  
140 145 150  
Asn Thr Pro Ala Phe Ser Leu Val Glu Asp Thr Asn Gly Asn Gln  
155 160 165  
Val Glu Arg Lys Ser Phe Asn Pro Trp Gly Leu Gln Cys His Gln  
170 175 180  
Ala His Leu Thr Arg Leu Cys Ser Glu Tyr Pro Glu Asn Lys Arg  
185 190 195  
His Arg Phe Leu Leu Leu Glu Asn Val Val Ser Gln Tyr Thr His  
200 205 210  
Pro Cys Val Leu Asp Leu Lys Met Gly Thr Arg Gln His Gly Asp  
215 220 225  
Asp Ala Ser Glu Glu Lys Lys Ala Arg His Met Arg Lys Cys Ala  
230 235 240  
Gln Ser Thr Ser Ala Cys Leu Gly Val Arg Ile Cys Gly Met Gln  
245 250 255  
Val Tyr Gln Thr Asp Lys Lys Tyr Phe Leu Cys Lys Asp Lys Tyr  
260 265 270  
Tyr Gly Arg Lys Leu Ser Val Glu Gly Phe Arg Gln Ala Leu Tyr  
275 280 285  
Gln Phe Leu His Asn Gly Ser His Leu Arg Arg Glu Leu Leu Glu  
290 295 300  
Pro Ile Leu His Gln Leu Arg Ala Leu Ser Ile Ile Arg Ser

305	310	315
Gln Ser Ser Tyr Arg Phe Tyr Ser Ser	Ser Leu Leu Val Ile Tyr	
320	325	330
Asp Gly Gln Glu Pro Pro Glu Arg Ala	Pro Gly Ser Pro His Pro	
335	340	345
His Glu Ala Pro Gln Ala Ala His Gly	Ser Ser Pro Gly Gly Leu	
350	355	360
Thr Lys Val Asp Ile Arg Met Ile Asp	Phe Ala His Thr Thr Tyr	
365	370	375
Lys Gly Tyr Trp Asn Glu His Thr Thr	Tyr Asp Gly Pro Asp Pro	
380	385	390
Gly Tyr Ile Phe Gly Leu Glu Asn Leu	Ile Arg Ile Leu Gln Asp	
395	400	405
Ile Gln Glu Gly Glu		
410		

&lt;210&gt; 16

&lt;211&gt; 1581

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477139CD1

&lt;400&gt; 16

Met Ala Gly Pro Gly Gly Trp Arg Asp Arg	Glu Val Thr Asp Leu		
1	5	10	15
Gly His Leu Pro Asp Pro Thr Gly Ile	Phe Ser Leu Asp Lys Thr		
20		25	30
Ile Gly Leu Gly Thr Tyr Gly Arg Ile	Tyr Leu Gly Leu His Glu		
35		40	45
Lys Thr Gly Ala Phe Thr Ala Val Lys	Val Met Asn Ala Arg Lys		
50		55	60
Thr Pro Leu Pro Glu Ile Gly Arg Arg	Val Arg Val Asn Lys Tyr		
65		70	75
Gln Lys Ser Val Gly Trp Arg Tyr Ser	Asp Glu Glu Asp Leu		
80		85	90
Arg Thr Glu Leu Asn Leu Leu Arg Lys	Tyr Ser Phe His Lys Asn		
95		100	105
Ile Val Ser Phe Tyr Gly Ala Phe Phe	Lys Leu Ser Pro Pro Gly		
110		115	120
Gln Arg His Gln Leu Trp Met Val Met	Glu Leu Cys Ala Ala Gly		
125		130	135
Ser Val Thr Asp Val Val Arg Met Thr	Ser Asn Gln Ser Leu Lys		
140		145	150
Glu Asp Trp Ile Ala Tyr Ile Cys Arg	Glu Ile Leu Gln Gly Leu		
155		160	165
Ala His Leu His Ala His Arg Val Ile	His Arg Asp Ile Lys Gly		
170		175	180
Gln Asn Val Leu Leu Thr His Asn Ala	Glu Val Lys Leu Val Asp		
185		190	195
Phe Gly Val Ser Ala Gln Val Ser Arg	Thr Asn Gly Arg Arg Asn		
200		205	210
Ser Phe Ile Gly Thr Pro Tyr Trp Met Ala	Pro Glu Val Ile Asp		
215		220	225

Cys Asp Glu Asp Pro Arg Arg Ser Tyr Asp Tyr Arg Ser Asp Val  
 230 235 240  
 Trp Ser Val Gly Ile Thr Ala Ile Glu Met Ala Glu Gly Ala Pro  
 245 250 255  
 Pro Leu Cys Asn Leu Gln Pro Leu Glu Ala Leu Phe Val Ile Leu  
 260 265 270  
 Arg Glu Ser Ala Pro Thr Val Lys Ser Ser Gly Trp Ser Arg Lys  
 275 280 285  
 Phe His Asn Phe Met Glu Lys Cys Thr Ile Lys Asn Phe Leu Phe  
 290 295 300  
 Arg Pro Thr Ser Ala Asn Met Leu Gln His Pro Phe Val Arg Asp  
 305 310 315  
 Ile Lys Asn Glu Arg His Val Val Glu Ser Leu Thr Arg His Leu  
 320 325 330  
 Thr Gly Ile Ile Lys Lys Arg Gln Lys Lys Gly Ile Pro Leu Ile  
 335 340 345  
 Phe Glu Arg Glu Glu Ala Ile Lys Glu Gln Tyr Thr Val Arg Arg  
 350 355 360  
 Phe Arg Gly Pro Ser Cys Thr His Glu Leu Leu Arg Leu Pro Thr  
 365 370 375  
 Ser Ser Arg Cys Arg Pro Leu Arg Val Leu His Gly Glu Pro Ser  
 380 385 390  
 Gln Pro Arg Trp Leu Pro Asp Arg Glu Glu Pro Gln Val Gln Ala  
 395 400 405  
 Leu Gln Gln Leu Gln Gly Ala Ala Arg Val Phe Met Pro Leu Gln  
 410 415 420  
 Ala Leu Asp Ser Ala Pro Lys Pro Leu Lys Gly Gln Ala Gln Ala  
 425 430 435  
 Pro Gln Arg Leu Gln Gly Ala Ala Arg Val Phe Met Pro Leu Gln  
 440 445 450  
 Ala Gln Val Lys Ala Lys Ala Ser Lys Pro Leu Gln Met Gln Ile  
 455 460 465  
 Lys Ala Pro Pro Arg Leu Arg Arg Ala Ala Arg Val Leu Met Pro  
 470 475 480  
 Leu Gln Ala Gln Val Arg Ala Pro Arg Leu Leu Gln Val Gln Ser  
 485 490 495  
 Gln Val Ser Lys Lys Gln Gln Ala Gln Thr Gln Thr Ser Glu Pro  
 500 505 510  
 Gln Asp Leu Asp Gln Val Pro Glu Glu Phe Gln Gly Gln Asp Gln  
 515 520 525  
 Val Pro Glu Gln Gln Arg Gln Gly Gln Ala Pro Glu Gln Gln Gln  
 530 535 540  
 Arg His Asn Gln Val Pro Glu Gln Glu Leu Glu Gln Asn Gln Ala  
 545 550 555  
 Pro Glu Gln Pro Glu Val Gln Glu Gln Ala Ala Glu Pro Ala Gln  
 560 565 570  
 Ala Glu Thr Glu Ala Glu Glu Pro Glu Ser Leu Arg Val Asn Ala  
 575 580 585  
 Gln Val Phe Leu Pro Leu Leu Ser Gln Asp His His Val Leu Leu  
 590 595 600  
 Pro Leu His Leu Asp Thr Gln Val Leu Ile Pro Val Glu Gly Gln  
 605 610 615  
 Thr Glu Gly Ser Pro Gln Ala Gln Ala Trp Thr Leu Glu Pro Pro  
 620 625 630  
 Gln Ala Ile Gly Ser Val Gln Ala Leu Ile Glu Gly Leu Ser Arg  
 635 640 645

Asp	Leu	Leu	Arg	Ala	Pro	Asn	Ser	Asn	Asn	Ser	Lys	Pro	Leu	Gly
650						655					660			
Pro	Leu	Gln	Thr	Leu	Met	Glu	Asn	Leu	Ser	Ser	Asn	Arg	Phe	Tyr
665						670					675			
Ser	Gln	Pro	Glu	Gln	Ala	Arg	Glu	Lys	Lys	Ser	Lys	Val	Ser	Thr
680						685					690			
Leu	Arg	Gln	Ala	Leu	Ala	Lys	Arg	Leu	Ser	Pro	Lys	Arg	Phe	Arg
695						700					705			
Ala	Lys	Ser	Ser	Trp	Arg	Pro	Glu	Lys	Leu	Glu	Leu	Ser	Asp	Leu
710						715					720			
Glu	Ala	Arg	Arg	Gln	Arg	Arg	Gln	Arg	Arg	Trp	Glu	Asp	Ile	Phe
725						730					735			
Asn	Gln	His	Glu	Glu	Glu	Leu	Arg	Gln	Val	Asp	Lys	Thr	Ser	Trp
740						745					750			
Arg	Gln	Trp	Gly	Pro	Ser	Asp	Gln	Leu	Ile	Asp	Asn	Ser	Phe	Thr
755						760					765			
Gly	Met	Gln	Asp	Leu	Lys	Lys	Tyr	Leu	Lys	Gly	Lys	Thr	Thr	Phe
770						775					780			
His	Asn	Val	Gln	Val	Val	Ile	Tyr	Arg	Ala	Val	Lys	Gly	Asn	Asp
785						790					795			
Asp	Val	Ala	Thr	Arg	Ser	Thr	Val	Pro	Gln	Arg	Ser	Leu	Glu	
800						805					810			
Gln	Ala	Gln	Lys	Pro	Ile	Asp	Ile	Arg	Gln	Arg	Ser	Ser	Gln	Asn
815						820					825			
Arg	Gln	Asn	Trp	Leu	Ala	Ala	Ser	Gly	Asp	Ser	Lys	His	Lys	Ile
830						835					840			
Leu	Ala	Gly	Lys	Thr	Gln	Ser	Tyr	Cys	Leu	Thr	Ile	Tyr	Ile	Ser
845						850					855			
Glu	Val	Lys	Lys	Glu	Glu	Phe	Gln	Glu	Gly	Met	Asn	Gln	Lys	Cys
860						865					870			
Gln	Gly	Ala	Gln	Val	Gly	Leu	Gly	Pro	Glu	Gly	His	Cys	Ile	Trp
875						880					885			
Gln	Leu	Gly	Glu	Ser	Ser	Ser	Glu	Glu	Ser	Pro	Val	Thr	Gly	
890						895					900			
Arg	Arg	Ser	Gln	Ser	Ser	Pro	Pro	Tyr	Ser	Thr	Ile	Asp	Gln	Lys
905						910					915			
Leu	Leu	Val	Asp	Ile	His	Val	Pro	Asp	Gly	Phe	Lys	Val	Gly	Lys
920						925					930			
Ile	Ser	Pro	Pro	Val	Tyr	Leu	Thr	Asn	Glu	Trp	Val	Gly	Tyr	Asn
935						940					945			
Ala	Leu	Ser	Glu	Ile	Phe	Arg	Asn	Asp	Trp	Leu	Thr	Pro	Ala	Pro
950						955					960			
Val	Ile	Gln	Pro	Pro	Glu	Glu	Asp	Gly	Asp	Tyr	Val	Glu	Leu	Tyr
965						970					975			
Asp	Ala	Ser	Ala	Asp	Thr	Asp	Gly	Asp	Asp	Asp	Asp	Glu	Ser	Asn
980						985					990			
Asp	Thr	Phe	Glu	Asp	Thr	Tyr	Asp	His	Ala	Asn	Gly	Asn	Asp	Asp
995						1000					1005			
Leu	Asp	Asn	Gln	Val	Asp	Gln	Ala	Asn	Asp	Val	Cys	Lys	Asp	His
1010						1015					1020			
Asp	Asp	Asp	Asn	Asn	Lys	Phe	Val	Asp	Asp	Val	Asn	Asn	Tyr	
1025						1030					1035			
Tyr	Glu	Ala	Pro	Ser	Cys	Pro	Ser	Leu	Leu	Ser	Gly	Gln	Ala	Met
1040						1045					1050			
Ala	Glu	Met	Glu	Ala	Ala	Ser	Lys	Met	Val	Met	Met	Glu	Val	Val
1055						1060					1065			

Glu Lys Arg Lys Pro Thr Glu Ala Met Glu Ala Ile Gln Pro Ile  
 1070 1075 1080  
 Glu Ala Met Glu Glu Val Gln Pro Val Arg Asp Asn Ala Ala Ile  
 1085 1090 1095  
 Gly Asp Gln Glu Glu His Ala Ala Asn Ile Gly Ser Glu Arg Arg  
 1100 1105 1110  
 Gly Ser Glu Gly Asp Gly Gly Lys Gly Val Val Arg Thr Ser Glu  
 1115 1120 1125  
 Glu Ser Gly Ala Leu Gly Leu Asn Gly Glu Glu Asn Cys Ser Glu  
 1130 1135 1140  
 Thr Asp Gly Pro Gly Leu Lys Arg Pro Ala Ser Gln Asp Phe Glu  
 1145 1150 1155  
 Tyr Leu Gln Glu Glu Pro Gly Gly Asn Glu Ala Ser Asn Ala  
 1160 1165 1170  
 Ile Asp Ser Gly Ala Ala Pro Ser Ala Pro Asp His Glu Ser Asp  
 1175 1180 1185  
 Asn Lys Asp Ile Ser Glu Ser Ser Thr Gln Ser Asp Phe Ser Ala  
 1190 1195 1200  
 Asn His Ser Ser Pro Ser Lys Gly Ser Gly Met Ser Ala Asp Ala  
 1205 1210 1215  
 Asn Phe Ala Ser Ala Ile Leu Tyr Ala Gly Phe Val Glu Val Pro  
 1220 1225 1230  
 Glu Glu Ser Pro Lys Gln Pro Ser Glu Val Asn Val Asn Pro Leu  
 1235 1240 1245  
 Tyr Val Ser Pro Ala Cys Lys Lys Pro Leu Ile His Met Tyr Glu  
 1250 1255 1260  
 Lys Glu Phe Thr Ser Glu Ile Cys Cys Gly Ser Leu Trp Gly Val  
 1265 1270 1275  
 Asn Leu Leu Leu Gly Thr Arg Ser Asn Leu Tyr Leu Met Asp Arg  
 1280 1285 1290  
 Ser Gly Lys Ala Asp Ile Thr Lys Leu Ile Arg Arg Arg Pro Phe  
 1295 1300 1305  
 Arg Gln Ile Gln Val Leu Glu Pro Leu Asn Leu Ile Thr Ile  
 1310 1315 1320  
 Ser Gly His Lys Asn Arg Leu Arg Val Tyr His Leu Thr Trp Leu  
 1325 1330 1335  
 Arg Asn Lys Ile Leu Asn Asn Asp Pro Glu Ser Lys Arg Arg Gln  
 1340 1345 1350  
 Glu Glu Met Leu Lys Thr Glu Glu Ala Cys Lys Ala Ile Asp Lys  
 1355 1360 1365  
 Leu Thr Gly Cys Glu His Phe Ser Val Leu Gln His Glu Glu Thr  
 1370 1375 1380  
 Thr Tyr Ile Ala Ile Ala Leu Lys Ser Ser Ile His Leu Tyr Ala  
 1385 1390 1395  
 Trp Ala Pro Lys Ser Phe Asp Glu Ser Thr Ala Ile Lys Val Phe  
 1400 1405 1410  
 Pro Thr Leu Asp His Lys Pro Val Thr Val Asp Leu Ala Ile Gly  
 1415 1420 1425  
 Ser Glu Lys Arg Leu Lys Ile Phe Phe Ser Ser Ala Asp Gly Tyr  
 1430 1435 1440  
 His Leu Ile Asp Ala Glu Ser Glu Val Met Ser Asp Val Thr Leu  
 1445 1450 1455  
 Pro Lys Asn Asn Ile Ile Leu Pro Asp Cys Leu Gly Ile Gly  
 1460 1465 1470  
 Met Met Leu Thr Phe Asn Ala Glu Ala Leu Ser Val Glu Ala Asn  
 1475 1480 1485

Glu Gln Leu Phe Lys Lys Ile Leu Glu Met Trp Lys Asp Ile Pro  
 1490 1495 1500  
 Ser Ser Ile Ala Phe Glu Cys Thr Gln Arg Thr Thr Gly Trp Gly  
 1505 1510 1515  
 Gln Lys Ala Ile Glu Val Arg Ser Leu Gln Ser Arg Val Leu Glu  
 1520 1525 1530  
 Ser Glu Leu Lys Arg Arg Ser Ile Lys Lys Leu Arg Phe Leu Cys  
 1535 1540 1545  
 Thr Arg Gly Asp Lys Leu Phe Phe Thr Ser Thr Leu Arg Asn His  
 1550 1555 1560  
 His Ser Arg Val Tyr Phe Met Thr Leu Gly Lys Leu Glu Glu Leu  
 1565 1570 1575  
 Gln Ser Asn Tyr Asp Val  
 1580

&lt;210&gt; 17

&lt;211&gt; 1084

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 55009053CD1

&lt;400&gt; 17

Met Glu Thr Gln Ala Val Ala Thr Ser Pro Asp Gly Arg Tyr Leu  
 1 5 10 15  
 Lys Phe Asp Ile Glu Ile Gly Arg Gly Ser Phe Lys Thr Val Tyr  
 20 25 30  
 Arg Gly Leu Asp Thr Asp Thr Thr Val Glu Val Ala Trp Cys Glu  
 35 40 45  
 Leu Gln Thr Arg Lys Leu Ser Arg Ala Glu Arg Gln Arg Phe Ser  
 50 55 60  
 Glu Glu Val Glu Met Leu Lys Gly Leu Gln His Pro Asn Ile Val  
 65 70 75  
 Arg Phe Tyr Asp Ser Trp Lys Ser Val Leu Arg Gly Gln Val Cys  
 80 85 90  
 Ile Val Leu Val Thr Glu Leu Met Thr Ser Gly Thr Leu Lys Thr  
 95 100 105  
 Tyr Leu Arg Arg Phe Arg Glu Met Lys Pro Arg Val Leu Gln Arg  
 110 115 120  
 Trp Ser Arg Gln Ile Leu Arg Gly Leu His Phe Leu His Ser Arg  
 125 130 135  
 Val Pro Pro Ile Leu His Arg Asp Leu Lys Cys Asp Asn Val Phe  
 140 145 150  
 Ile Thr Gly Pro Ser Gly Ser Val Lys Ile Gly Asp Leu Gly Leu  
 155 160 165  
 Ala Thr Leu Lys Arg Ala Ser Phe Ala Lys Ser Val Ile Gly Thr  
 170 175 180  
 Pro Glu Phe Met Ala Pro Glu Met Tyr Glu Glu Lys Tyr Asp Glu  
 185 190 195  
 Ala Val Asp Val Tyr Ala Phe Gly Met Cys Met Leu Glu Met Ala  
 200 205 210  
 Thr Ser Glu Tyr Pro Tyr Ser Glu Cys Gln Asn Ala Ala Gln Ile  
 215 220 225  
 Tyr Arg Lys Val Thr Ser Gly Arg Lys Pro Asn Ser Phe His Lys

230	235	240
Val Lys Ile Pro Glu Val Lys Glu Ile Ile	Glu Gly Cys Ile Arg	
245	250	255
Thr Asp Lys Asn Glu Arg Phe Thr Ile Gln	Asp Leu Leu Ala His	
260	265	270
Ala Phe Phe Arg Glu Glu Arg Gly Val His	Val Glu Leu Ala Glu	
275	280	285
Glu Asp Asp Gly Glu Lys Pro Gly Leu Lys	Leu Trp Leu Arg Met	
290	295	300
Glu Asp Ala Arg Arg Gly Arg Pro Arg Asp	Asn Gln Ala Ile	
305	310	315
Glu Phe Leu Phe Gln Leu Gly Arg Asp Ala	Ala Glu Glu Val Ala	
320	325	330
Gln Glu Met Val Ala Leu Gly Leu Val Cys	Glu Ala Asp Tyr Gln	
335	340	345
Pro Val Ala Arg Ala Val Arg Glu Arg Val	Ala Ala Ile Gln Arg	
350	355	360
Lys Arg Glu Lys Leu Arg Lys Ala Arg Glu	Leu Glu Ala Leu Pro	
365	370	375
Pro Glu Pro Gly Pro Pro Pro Ala Thr Val	Pro Met Ala Pro Gly	
380	385	390
Pro Pro Ser Val Phe Pro Pro Glu Pro Glu	Pro Glu Ala Asp	
395	400	405
Gln His Gln Pro Phe Leu Phe Arg His Ala	Ser Tyr Ser Ser Thr	
410	415	420
Thr Ser Asp Cys Glu Thr Asp Gly Tyr Leu	Ser Ser Ser Gly Phe	
425	430	435
Leu Asp Ala Ser Asp Pro Ala Leu Gln Pro	Pro Pro Gly Gly Val Pro	
440	445	450
Ser Ser Leu Ala Glu Ser His Leu Cys Leu	Pro Ser Ala Phe Ala	
455	460	465
Leu Ser Ile Pro Arg Ser Gly Pro Gly Ser	Asp Phe Ser Pro Gly	
470	475	480
Asp Ser Tyr Ala Ser Asp Ala Ala Ser Gly	Leu Ser Asp Val Gly	
485	490	495
Glu Gly Met Gly Gln Met Arg Arg Pro Pro	Gly Arg Asn Leu Arg	
500	505	510
Arg Arg Pro Arg Ser Arg Leu Arg Val Thr	Ser Val Asp Gln	
515	520	525
Asn Asp Arg Val Val Glu Cys Gln Leu Gln	Thr His Asn Ser Lys	
530	535	540
Met Val Thr Phe Arg Phe Asp Leu Asp Gly	Asp Ser Pro Glu Glu	
545	550	555
Ile Ala Ala Ala Met Val Tyr Asn Glu Phe	Ile Leu Pro Ser Glu	
560	565	570
Arg Asp Gly Phe Leu Arg Arg Ile Arg Glu	Ile Ile Gln Arg Val	
575	580	585
Glu Thr Leu Leu Lys Arg Asp Thr Gly Pro	Met Glu Ala Ala Glu	
590	595	600
Asp Thr Leu Ser Pro Gln Glu Glu Pro Ala	Pro Leu Pro Ala Leu	
605	610	615
Pro Val Pro Leu Pro Asp Pro Ser Asn Glu	Glu Leu Gln Ser Ser	
620	625	630
Thr Ser Leu Glu His Arg Ser Trp Thr Ala	Phe Ser Thr Ser Ser	
635	640	645
Ser Ser Pro Gly Thr Pro Leu Ser Pro Gly	Asn Pro Phe Ser Pro	

650	655	660
Gly Thr Pro Ile Ser Pro Gly Pro Ile Phe Pro Ile Thr Ser Pro		
665	670	675
Pro Cys His Pro Ser Pro Ser Pro Phe Ser Pro Ile Ser Ser Gln		
680	685	690
Val Ser Ser Asn Pro Ser Pro His Pro Thr Ser Ser Pro Leu Pro		
695	700	705
Phe Ser Ser Ser Thr Pro Glu Phe Pro Val Pro Leu Ser Gln Cys		
710	715	720
Pro Trp Ser Ser Leu Pro Thr Thr Ser Pro Pro Thr Phe Ser Pro		
725	730	735
Thr Cys Ser Gln Val Thr Leu Ser Ser Pro Phe Phe Pro Pro Cys		
740	745	750
Pro Ser Thr Ser Ser Phe Pro Ser Thr Thr Ala Ala Pro Leu Leu		
755	760	765
Ser Leu Ala Ser Ala Phe Ser Leu Ala Val Met Thr Val Ala Gln		
770	775	780
Ser Leu Leu Ser Pro Ser Pro Gly Leu Leu Ser Gln Ser Pro Pro		
785	790	795
Ala Pro Pro Ser Pro Leu Pro Ser Leu Pro Leu Pro Pro Pro Val		
800	805	810
Ala Pro Gly Gly Gln Glu Ser Pro Ser Pro His Thr Ala Glu Val		
815	820	825
Glu Ser Glu Ala Ser Pro Pro Pro Ala Arg Pro Leu Pro Gly Glu		
830	835	840
Ala Arg Leu Ala Pro Ile Ser Glu Glu Gly Lys Pro Gln Leu Val		
845	850	855
Gly Arg Phe Gln Val Thr Ser Ser Lys Glu Pro Ala Glu Pro Leu		
860	865	870
Pro Leu Gln Pro Thr Ser Pro Thr Leu Ser Gly Ser Pro Lys Pro		
875	880	885
Ser Thr Pro Gln Leu Thr Ser Glu Ser Ser Asp Thr Glu Asp Ser		
890	895	900
Ala Gly Gly Pro Glu Thr Arg Glu Ala Leu Ala Glu Ser Asp		
905	910	915
Arg Ala Ala Glu Gly Leu Gly Ala Gly Val Glu Glu Glu Gly Asp		
920	925	930
Asp Gly Lys Glu Pro Gln Val Gly Gly Ser Pro Gln Pro Leu Ser		
935	940	945
His Pro Ser Pro Val Trp Met Asn Tyr Ser Tyr Ser Ser Leu Cys		
950	955	960
Leu Ser Ser Glu Glu Ser Glu Ser Ser Gly Glu Asp Glu Glu Phe		
965	970	975
Trp Ala Glu Leu Gln Ser Leu Arg Gln Lys His Leu Ser Glu Val		
980	985	990
Glu Thr Leu Gln Thr Leu Gln Lys Lys Glu Ile Glu Asp Leu Tyr		
995	1000	1005
Ser Arg Leu Gly Lys Gln Pro Pro Pro Gly Ile Val Ala Pro Ala		
1010	1015	1020
Ala Met Leu Ser Ser Arg Gln Arg Arg Leu Ser Lys Gly Ser Phe		
1025	1030	1035
Pro Thr Ser Arg Arg Asn Ser Leu Gln Arg Ser Glu Pro Pro Gly		
1040	1045	1050
Pro Gly Ile Met Arg Arg Asn Ser Leu Ser Gly Ser Ser Thr Gly		
1055	1060	1065
Ser Gln Glu Gln Arg Ala Ser Lys Gly Val Thr Phe Ala Gly Asp		

1070

1075

1080

Val Gly Arg Met

&lt;210&gt; 18

&lt;211&gt; 600

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474648CD1

&lt;400&gt; 18

Met	Gly	Glu	Ser	Gly	Asn	His	His	Phe	Gln	Gln	Thr	Asn	Thr	Gly
1				5					10			15		
Thr	Glu	Asn	Gln	Thr	Ala	His	Val	Leu	Thr	His	Lys	Trp	Glu	Leu
				20					25			30		
Asp	Asn	Glu	Asn	Ile	Trp	Ala	Gln	Gly	Gly	Glu	His	His	Lys	Leu
				35					40			45		
Gly	Pro	Val	Met	Gly	Trp	Ala	Arg	Ser	Gly	Lys	Thr	Leu	Gly	
				50					55			60		
Glu	Ile	Pro	Asn	Val	Gly	Thr	Leu	Thr	Leu	Leu	Thr	Gly	Tyr	Gly
				65					70			75		
Gly	Cys	Gln	Leu	Pro	Cys	Cys	Lys	Asp	Thr	Gln	Ala	Ala	Tyr	Gly
				80					85			90		
Glu	Thr	His	Val	Val	Arg	Ser	Gly	Gly	Leu	Leu	Pro	Thr	Ala	Ser
				95					100			105		
Trp	Glu	Leu	Arg	Pro	Ala	Asp	Ser	His	Thr	Val	Thr	Ser	Asp	Asp
				110					115			120		
Pro	Gly	Val	Ser	Val	Val	Ser	Gly	Tyr	Pro	Gly	Gly	Cys	Leu	Pro
				125					130			135		
Asp	His	Asp	Pro	Pro	Val	Gly	Phe	Leu	Ser	Glu	Gly	Pro	Ala	Pro
				140					145			150		
Arg	Ser	Cys	Ser	Leu	Ile	Lys	Gly	Gly	Gly	Thr	Gly	Leu	Ala	Ala
				155					160			165		
Ser	Arg	Val	Pro	Arg	Ser	Arg	Glu	Arg	Arg	Ala	Cys	Cys	Gly	Tyr
				170					175			180		
Gly	Val	Arg	Arg	Gln	Gln	Glu	Gly	Gly	Pro	Gly	Ala	Thr	Ser	Ala
				185					190			195		
Gly	Leu	Gly	Gln	Ala	Arg	Arg	Ser	Lys	Pro	Ser	Arg	Arg	Arg	Arg
				200					205			210		
Arg	Gly	Ala	Trp	Ala	Arg	Gly	Gly	Pro	Gly	Gly	Ala	Glu	Asp	
				215					220			225		
Thr	Gly	Gly	Ser	Leu	Pro	Ser	Gln	Val	Arg	Pro	Pro	Gly	Pro	Cys
				230					235			240		
Gln	Cys	Pro	Val	Gln	Phe	Leu	Phe	Asp	Ile	Ser	Glu	Gln	Gly	Val
				245					250			255		
Gln	Arg	Met	Gly	Lys	Lys	Arg	Ala	Gly	Ala	Ala	Asn	Lys	Gly	
				260					265			270		
Arg	Asn	Ser	Tyr	Leu	Arg	Arg	Tyr	Asp	Ile	Lys	Ala	Leu	Ile	Gly
				275					280			285		
Thr	Gly	Ser	Phe	Ser	Arg	Val	Val	Arg	Val	Glu	Gln	Lys	Thr	Thr
				290					295			300		
Lys	Lys	Pro	Phe	Ala	Ile	Lys	Val	Met	Glu	Thr	Arg	Glu	Arg	Glu
				305					310			315		

Gly	Arg	Glu	Ala	Cys	Val	Ser	Glu	Leu	Ser	Val	Leu	Arg	Arg	Val
					320			325				330		
Ser	His	Arg	Tyr	Ile	Val	Gln	Leu	Met	Glu	Ile	Phe	Glu	Thr	Glu
					335			340				345		
Asp	Gln	Val	Tyr	Met	Val	Met	Glu	Leu	Ala	Thr	Gly	Gly	Glu	Leu
					350			355				360		
Phe	Asp	Arg	Leu	Ile	Ala	Gln	Gly	Ser	Phe	Thr	Glu	Arg	Asp	Ala
					365			370				375		
Val	Arg	Ile	Leu	Gln	Met	Val	Ala	Asp	Gly	Ile	Arg	Tyr	Leu	His
					380			385				390		
Ala	Leu	Gln	Ile	Thr	His	Arg	Asn	Leu	Lys	Pro	Glu	Asn	Leu	Leu
					395			400				405		
Tyr	Tyr	His	Pro	Gly	Glu	Glu	Ser	Lys	Ile	Leu	Ile	Thr	Asp	Phe
					410			415				420		
Gly	Leu	Ala	Tyr	Ser	Gly	Lys	Lys	Ser	Gly	Asp	Trp	Thr	Met	Lys
					425			430				435		
Thr	Leu	Cys	Gly	Thr	Pro	Glu	Tyr	Ile	Ala	Pro	Glu	Val	Leu	Leu
					440			445				450		
Arg	Lys	Pro	Tyr	Thr	Ser	Ala	Val	Asp	Met	Trp	Ala	Leu	Gly	Val
					455			460				465		
Ile	Thr	Tyr	Ala	Leu	Leu	Ser	Gly	Phe	Leu	Pro	Phe	Asp	Asp	Glu
					470			475				480		
Ser	Gln	Thr	Arg	Leu	Tyr	Arg	Lys	Ile	Leu	Lys	Gly	Lys	Tyr	Asn
					485			490				495		
Tyr	Thr	Gly	Glu	Pro	Trp	Pro	Ser	Ile	Ser	His	Leu	Ala	Lys	Asp
					500			505				510		
Phe	Ile	Asp	Lys	Leu	Leu	Ile	Leu	Glu	Ala	Gly	His	Arg	Met	Ser
					515			520				525		
Ala	Gly	Gln	Ala	Leu	Asp	His	Pro	Trp	Val	Ile	Thr	Met	Ala	Ala
					530			535				540		
Gly	Ser	Ser	Met	Lys	Asn	Leu	Gln	Arg	Ala	Ile	Ser	Arg	Asn	Leu
					545			550				555		
Met	Gln	Arg	Ala	Ser	Pro	His	Ser	Gln	Ser	Pro	Gly	Ser	Ala	Gln
					560			565				570		
Ser	Ser	Lys	Ser	His	Tyr	Ser	His	Lys	Ser	Arg	His	Met	Trp	Ser
					575			580				585		
Lys	Arg	Asn	Leu	Arg	Ile	Val	Glu	Ser	Pro	Leu	Ser	Ala	Leu	Leu
					590			595				600		

<210> 19  
 <211> 1114  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7483053CD1

<400> 19  
 Met Ala Lys Ala Thr Ser Gly Ala Ala Gly Leu Arg Leu Leu Leu  
 1 5 10 15  
 Leu Leu Leu Pro Leu Leu Gly Lys Val Ala Leu Gly Leu Tyr  
 20 25 30  
 Phe Ser Arg Asp Ala Tyr Trp Glu Lys Leu Tyr Val Asp Gln Ala  
 35 40 45

Ala Gly Thr Pro Leu Leu Tyr Val His Ala Leu Arg Asp Ala Pro  
                   50                  55                  60  
 Glu Glu Val Pro Ser Phe Arg Leu Gly Gln His Leu Tyr Gly Thr  
                   65                  70                  75  
 Tyr Arg Thr Arg Leu His Glu Asn Asn Trp Ile Cys Ile Gln Glu  
                   80                  85                  90  
 Asp Thr Gly Leu Leu Tyr Leu Asn Arg Ser Leu Asp His Ser Ser  
                   95                  100                 105  
 Trp Glu Lys Leu Ser Val Arg Asn Arg Gly Phe Pro Leu Leu Thr  
                   110                  115                 120  
 Val Tyr Leu Lys Val Phe Leu Ser Pro Thr Ser Leu Arg Glu Gly  
                   125                  130                 135  
 Glu Cys Gln Trp Pro Gly Cys Ala Arg Val Tyr Phe Ser Phe Phe  
                   140                  145                 150  
 Asn Thr Ser Phe Pro Ala Cys Ser Ser Leu Lys Pro Arg Glu Leu  
                   155                  160                 165  
 Cys Phe Pro Glu Thr Arg Pro Ser Phe Arg Ile Arg Glu Asn Arg  
                   170                  175                 180  
 Pro Pro Gly Thr Phe His Gln Phe Arg Leu Leu Pro Val Gln Phe  
                   185                  190                 195  
 Leu Cys Pro Asn Ile Ser Val Ala Tyr Arg Leu Leu Glu Gly Glu  
                   200                  205                 210  
 Gly Leu Pro Phe Arg Cys Ala Pro Asp Ser Leu Glu Val Ser Thr  
                   215                  220                 225  
 Arg Trp Ala Leu Asp Arg Glu Gln Arg Glu Lys Tyr Glu Leu Val  
                   230                  235                 240  
 Ala Val Cys Thr Val His Ala Gly Ala Arg Glu Glu Val Val Met  
                   245                  250                 255  
 Val Pro Phe Pro Val Thr Val Tyr Asp Glu Asp Asp Ser Ala Pro  
                   260                  265                 270  
 Thr Phe Pro Ala Gly Val Asp Thr Ala Ser Ala Val Val Glu Phe  
                   275                  280                 285  
 Lys Arg Lys Glu Asp Thr Val Val Ala Thr Leu Arg Val Phe Asp  
                   290                  295                 300  
 Ala Asp Val Val Pro Ala Ser Gly Glu Leu Val Arg Arg Tyr Thr  
                   305                  310                 315  
 Ser Thr Leu Leu Pro Gly Asp Thr Trp Ala Gln Gln Thr Phe Arg  
                   320                  325                 330  
 Val Glu His Trp Pro Asn Glu Thr Ser Val Gln Ala Asn Gly Ser  
                   335                  340                 345  
 Phe Val Arg Ala Thr Val His Asp Tyr Arg Leu Val Leu Asn Arg  
                   350                  355                 360  
 Asn Leu Ser Ile Ser Glu Asn Arg Thr Met Gln Leu Ala Val Leu  
                   365                  370                 375  
 Val Asn Asp Ser Asp Phe Gln Gly Pro Gly Ala Gly Val Leu Leu  
                   380                  385                 390  
 Leu His Phe Asn Val Ser Val Leu Pro Val Ser Leu His Leu Pro  
                   395                  400                 405  
 Ser Thr Tyr Ser Leu Ser Val Ser Arg Arg Ala Arg Arg Phe Ala  
                   410                  415                 420  
 Gln Ile Gly Lys Val Cys Val Glu Asn Cys Gln Ala Phe Ser Gly  
                   425                  430                 435  
 Ile Asn Val Gln Tyr Lys Leu His Ser Ser Gly Ala Asn Cys Ser  
                   440                  445                 450  
 Thr Leu Gly Val Val Thr Ser Ala Glu Asp Thr Ser Gly Ile Leu  
                   455                  460                 465

Phe Val Asn Asp Thr Lys Ala Leu Arg Arg Pro Lys Cys Ala Glu  
 470 475 480  
 Leu His Tyr Met Val Val Ala Thr Asp Gln Gln Thr Ser Arg Gln  
 485 490 495  
 Ala Gln Ala Gln Leu Leu Val Thr Val Glu Gly Ser Tyr Val Ala  
 500 505 510  
 Glu Glu Ala Gly Cys Pro Leu Ser Cys Ala Val Ser Lys Arg Arg  
 515 520 525  
 Leu Glu Cys Glu Glu Cys Gly Gly Leu Gly Ser Pro Thr Gly Arg  
 530 535 540  
 Cys Glu Trp Arg Gln Gly Asp Gly Lys Gly Ile Thr Arg Asn Phe  
 545 550 555  
 Ser Thr Cys Ser Pro Ser Thr Lys Thr Cys Pro Asp Gly His Cys  
 560 565 570  
 Asp Val Val Glu Thr Gln Asp Ile Asn Ile Cys Pro Gln Asp Cys  
 575 580 585  
 Leu Arg Gly Ser Ile Val Gly Gly His Glu Pro Gly Glu Pro Arg  
 590 595 600  
 Gly Ile Lys Ala Gly Tyr Gly Thr Cys Asn Cys Phe Pro Glu Glu  
 605 610 615  
 Glu Lys Cys Phe Cys Glu Pro Glu Asp Ile Gln Asp Pro Leu Cys  
 620 625 630  
 Asp Glu Leu Cys Arg Thr Val Ile Ala Ala Ala Val Leu Phe Ser  
 635 640 645  
 Phe Ile Val Ser Val Leu Leu Ser Ala Phe Cys Ile His Cys Tyr  
 650 655 660  
 His Lys Phe Ala His Lys Pro Pro Ile Ser Ser Ala Glu Met Thr  
 665 670 675  
 Phe Arg Arg Pro Ala Gln Ala Phe Pro Val Ser Tyr Ser Ser Ser  
 680 685 690  
 Ser Ala Arg Arg Pro Ser Leu Asp Ser Met Glu Asn Gln Val Ser  
 695 700 705  
 Val Asp Ala Phe Lys Ile Leu Glu Asp Pro Lys Trp Glu Phe Pro  
 710 715 720  
 Arg Lys Asn Leu Val Leu Gly Lys Thr Leu Gly Glu Gly Glu Phe  
 725 730 735  
 Gly Lys Val Val Lys Ala Thr Ala Phe His Leu Lys Gly Arg Ala  
 740 745 750  
 Gly Tyr Thr Thr Val Ala Val Lys Met Leu Lys Glu Asn Ala Ser  
 755 760 765  
 Pro Ser Glu Leu Arg Asp Leu Leu Ser Glu Phe Asn Val Leu Lys  
 770 775 780  
 Gln Val Asn His Pro His Val Ile Lys Leu Tyr Gly Ala Cys Ser  
 785 790 795  
 Gln Asp Gly Pro Leu Leu Leu Ile Val Glu Tyr Ala Lys Tyr Gly  
 800 805 810  
 Ser Leu Arg Gly Phe Leu Arg Glu Ser Arg Lys Val Gly Pro Gly  
 815 820 825  
 Tyr Leu Gly Ser Gly Gly Ser Arg Asn Ser Ser Ser Leu Asp His  
 830 835 840  
 Pro Asp Glu Arg Ala Leu Thr Met Gly Asp Leu Ile Ser Phe Ala  
 845 850 855  
 Trp Gln Ile Ser Gln Gly Met Gln Tyr Leu Ala Glu Met Lys Leu  
 860 865 870  
 Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Ala Glu Gly  
 875 880 885

Arg Lys Met Lys Ile Ser Asp Phe Gly Leu Ser Arg Asp Val Tyr  
 890 895 900  
 Glu Glu Asp Ser Tyr Val Lys Arg Ser Gln Gly Arg Ile Pro Val  
 905 910 915  
 Lys Trp Met Ala Ile Glu Ser Leu Phe Asp His Ile Tyr Thr Thr  
 920 925 930  
 Gln Ser Asp Val Trp Ser Phe Gly Val Leu Leu Trp Glu Ile Val  
 935 940 945  
 Thr Leu Gly Gly Asn Pro Tyr Pro Gly Ile Pro Pro Glu Arg Leu  
 950 955 960  
 Phe Asn Leu Leu Lys Thr Gly His Arg Met Glu Arg Pro Asp Asn  
 965 970 975  
 Cys Ser Glu Glu Met Tyr Arg Leu Met Leu Gln Cys Trp Lys Gln  
 980 985 990  
 Glu Pro Asp Lys Arg Pro Val Phe Ala Asp Ile Ser Lys Asp Leu  
 995 1000 1005  
 Glu Lys Met Met Val Lys Arg Arg Asp Tyr Leu Asp Leu Ala Ala  
 1010 1015 1020  
 Ser Thr Pro Ser Asp Ser Leu Ile Tyr Asp Asp Gly Leu Ser Glu  
 1025 1030 1035  
 Glu Glu Thr Pro Leu Val Asp Cys Asn Asn Ala Pro Leu Pro Arg  
 1040 1045 1050  
 Ala Leu Pro Ser Thr Trp Ile Glu Asn Lys Leu Tyr Gly Met Ser  
 1055 1060 1065  
 Asp Pro Asn Trp Pro Gly Glu Ser Pro Val Pro Leu Thr Arg Ala  
 1070 1075 1080  
 Asp Gly Thr Asn Thr Gly Phe Pro Arg Tyr Pro Asn Asp Ser Val  
 1085 1090 1095  
 Tyr Ala Asn Trp Met Leu Ser Pro Ser Ala Ala Lys Leu Met Asp  
 1100 1105 1110  
 Thr Phe Asp Ser

<210> 20  
 <211> 567  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7483117CD1

<400> 20  
 Met Asp Asp Lys Asp Ile Asp Lys Glu Leu Arg Gln Lys Leu Asn  
 1 5 10 15  
 Phe Ser Tyr Cys Glu Glu Thr Glu Ile Glu Gly Gln Lys Lys Val  
 20 25 30  
 Glu Glu Ser Arg Glu Ala Ser Ser Gln Thr Pro Glu Lys Gly Glu  
 35 40 45  
 Val Gln Asp Ser Glu Ala Lys Gly Thr Pro Pro Trp Thr Pro Leu  
 50 55 60  
 Ser Asn Val His Glu Leu Asp Thr Ser Ser Glu Lys Asp Lys Glu  
 65 70 75  
 Ser Pro Asp Gln Ile Leu Arg Thr Pro Val Ser His Pro Leu Lys  
 80 85 90  
 Cys Pro Glu Thr Pro Ala Gln Pro Asp Ser Arg Ser Lys Leu Leu

95	100	105
Pro Ser Asp Ser	Pro Ser Thr Pro Lys	Met Leu Ser Arg
110	115	120
Val Ile Ser Pro Thr Gly Lys	Leu Pro Ser Arg Gly	Pro Lys His
125	130	135
Leu Lys Leu Thr Pro Ala Pro	Leu Lys Asp Glu Met	Thr Ser Leu
140	145	150
Ala Leu Val Asn Ile Asn Pro	Phe Thr Pro Glu Ser	Tyr Lys Lys
155	160	165
Leu Phe Leu Gln Ser Gly Gly	Lys Arg Lys Ile Arg	Gly Asp Leu
170	175	180
Glu Glu Ala Gly Pro Glu Glu	Gly Lys Gly Leu Pro	Ala Lys
185	190	195
Arg Cys Val Leu Arg Glu Thr	Asn Met Ala Ser Arg	Tyr Glu Lys
200	205	210
Glu Phe Leu Glu Val Glu Lys	Ile Gly Val Gly Glu	Phe Gly Thr
215	220	225
Val Tyr Lys Cys Ile Lys Arg	Leu Asp Gly Cys Val	Tyr Ala Ile
230	235	240
Lys Arg Ser Met Lys Thr Phe	Thr Glu Leu Ser Asn	Glu Asn Ser
245	250	255
Ala Leu His Glu Val Tyr Ala	His Ala Val Leu Gly	His His Pro
260	265	270
His Val Val Arg Tyr Tyr Ser	Ser Trp Ala Glu Asp	Asp His Met
275	280	285
Ile Ile Gln Asn Glu Tyr Cys	Asn Gly Ser Leu Gln	Ala Ala
290	295	300
Ile Ser Glu Asn Thr Lys Ser	Gly Asn His Phe	Glu Glu Pro Lys
305	310	315
Leu Lys Asp Ile Leu Leu Gln	Ile Ser Leu Gly	Leu Asn Tyr Ile
320	325	330
His Asn Ser Ser Met Val His	Leu Asp Ile Lys Pro	Ser Asn Ile
335	340	345
Phe Ile Cys His Lys Met Gln	Ser Glu Ser Ser	Gly Val Ile Glu
350	355	360
Glu Val Glu Asn Glu Ala Asp	Trp Phe Leu Ser	Ala Asn Val Met
365	370	375
Tyr Lys Ile Gly Asp Leu Gly	His Ala Thr Ser Ile	Asn Lys Pro
380	385	390
Lys Val Glu Glu Gly Asp Ser	Arg Phe Leu Ala Asn	Glu Ile Leu
395	400	405
Gln Glu Asp Tyr Arg His	Leu Pro Lys Ala Asp	Ile Phe Ala Leu
410	415	420
Gly Leu Thr Ile Ala Val Ala	Ala Gly Ala Glu Ser	Leu Pro Thr
425	430	435
Asn Gly Ala Ala Trp His His	Ile Arg Lys Gly Asn	Phe Pro Asp
440	445	450
Val Pro Gln Glu Leu Ser	Glu Ser Phe Ser Ser	Leu Leu Lys Asn
455	460	465
Met Ile Gln Pro Asp Ala Glu	Gln Arg Pro Ser	Ala Ala Ala Leu
470	475	480
Ala Arg Asn Thr Val Leu Arg	Pro Ser Leu Gly	Lys Thr Glu Glu
485	490	495
Leu Gln Gln Gln Leu Asn Leu	Glu Lys Phe Lys	Thr Ala Thr Leu
500	505	510
Glu Arg Glu Leu Arg Glu Ala	Gln Gln Ala Gln Ser	Pro Gln Gly

515	520	525
Tyr Thr His His Gly Asp Thr Gly Val	Ser Gly Thr His Thr Gly	
530	535	540
Ser Arg Ser Thr Lys Arg Leu Val Gly	Gly Lys Ser Ala Arg Ser	
545	550	555
Ser Ser Phe Thr Ser Gly Glu Arg Glu	Pro Leu His	
560	565	

<210> 21

<211> 2054

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7484498CD1

<400> 21

Met Leu Lys Phe Lys Tyr Gly Ala Arg Asn Pro Leu Asp Ala Gly			
1	5	10	15
Ala Ala Glu Pro Ile Ala Ser Arg Ala Ser Arg Leu Asn Leu Phe			
20	25		30
Phe Gln Gly Lys Pro Pro Phe Met Thr Gln Gln Gln Met Ser Pro			
35	40		45
Leu Ser Arg Glu Gly Ile Leu Asp Ala Leu Phe Val Leu Phe Glu			
50	55		60
Glu Cys Ser Gln Pro Ala Leu Met Lys Ile Lys His Val Ser Asn			
65	70		75
Phe Val Arg Lys Tyr Ser Asp Thr Ile Ala Glu Leu Gln Glu Leu			
80	85		90
Gln Pro Ser Ala Lys Asp Phe Glu Val Arg Ser Leu Val Gly Cys			
95	100		105
Gly His Phe Ala Glu Val Gln Val Val Arg Glu Lys Ala Thr Gly			
110	115		120
Asp Ile Tyr Ala Met Lys Val Met Lys Lys Ala Leu Leu Ala			
125	130		135
Gln Glu Gln Val Ser Phe Phe Glu Glu Glu Arg Asn Ile Leu Ser			
140	145		150
Arg Ser Thr Ser Pro Trp Ile Pro Gln Leu Gln Tyr Ala Phe Gln			
155	160		165
Asp Lys Asn His Leu Tyr Leu Val Met Glu Tyr Gln Pro Gly Gly			
170	175		180
Asp Leu Leu Ser Leu Leu Asn Arg Tyr Glu Asp Gln Leu Asp Glu			
185	190		195
Asn Leu Ile Gln Phe Tyr Leu Ala Glu Leu Ile Leu Ala Val His			
200	205		210
Ser Val His Leu Met Gly Tyr Val His Arg Asp Ile Lys Pro Glu			
215	220		225
Asn Ile Leu Val Asp Arg Thr Gly His Ile Lys Leu Val Asp Phe			
230	235		240
Gly Ser Ala Ala Lys Met Asn Ser Asn Lys Met Val Asn Ala Lys			
245	250		255
Leu Pro Ile Gly Thr Pro Asp Tyr Met Ala Pro Glu Val Leu Thr			
260	265		270
Val Met Asn Gly Asp Gly Lys Gly Thr Tyr Gly Leu Asp Cys Asp			
275	280		285

Trp Trp Ser Val Gly Val Ile Ala Tyr Glu Met Ile Tyr Gly Arg  
290 295 300  
Ser Pro Phe Ala Glu Gly Thr Ser Ala Arg Thr Phe Asn Asn Ile  
305 310 315  
Met Asn Phe Gln Arg Phe Leu Lys Phe Pro Asp Asp Pro Lys Val  
320 325 330  
Ser Ser Asp Phe Leu Asp Leu Ile Gln Ser Leu Leu Cys Gly Gln  
335 340 345  
Lys Glu Arg Leu Lys Phe Glu Gly Leu Cys Cys His Pro Phe Phe  
350 355 360  
Ser Lys Ile Asp Trp Asn Asn Ile Arg Asn Ser Pro Pro Pro Phe  
365 370 375  
Val Pro Thr Leu Lys Ser Asp Asp Asp Thr Ser Asn Phe Asp Glu  
380 385 390  
Pro Glu Lys Asn Ser Trp Val Ser Ser Pro Cys Gln Leu Ser  
395 400 405  
Pro Ser Gly Phe Ser Gly Glu Glu Leu Pro Phe Val Gly Phe Ser  
410 415 420  
Tyr Ser Lys Ala Leu Gly Ile Leu Gly Arg Ser Glu Ser Val Val  
425 430 435  
Ser Gly Leu Asp Ser Pro Ala Lys Thr Ser Ser Met Glu Lys Lys  
440 445 450  
Leu Leu Ile Lys Ser Lys Glu Leu Gln Asp Ser Gln Asp Lys Cys  
455 460 465  
His Lys Met Glu Gln Glu Met Thr Arg Leu His Arg Arg Val Ser  
470 475 480  
Glu Val Glu Ala Val Leu Ser Gln Lys Glu Val Glu Leu Lys Ala  
485 490 495  
Ser Glu Thr Gln Arg Ser Leu Leu Glu Gln Asp Leu Ala Thr Tyr  
500 505 510  
Ile Thr Glu Cys Ser Ser Leu Lys Arg Ser Leu Glu Gln Ala Arg  
515 520 525  
Met Glu Val Ser Gln Glu Asp Asp Lys Ala Leu Gln Leu Leu His  
530 535 540  
Asp Ile Arg Glu Gln Ser Arg Lys Leu Gln Glu Ile Lys Glu Gln  
545 550 555  
Glu Tyr Gln Ala Gln Val Glu Glu Met Arg Leu Met Met Asn Gln  
560 565 570  
Leu Glu Glu Asp Leu Val Ser Ala Arg Arg Arg Ser Asp Leu Tyr  
575 580 585  
Glu Ser Glu Leu Arg Glu Ser Arg Leu Ala Ala Glu Glu Phe Lys  
590 595 600  
Arg Lys Ala Thr Glu Cys Gln His Lys Leu Leu Lys Ala Lys Asp  
605 610 615  
Gln Gly Lys Pro Glu Val Gly Glu Tyr Ala Lys Leu Glu Lys Ile  
620 625 630  
Asn Ala Glu Gln Gln Leu Lys Ile Gln Glu Leu Gln Glu Lys Leu  
635 640 645  
Glu Lys Ala Val Lys Ala Ser Thr Glu Ala Thr Glu Leu Leu Gln  
650 655 660  
Asn Ile Arg Gln Ala Lys Glu Arg Ala Glu Arg Glu Leu Glu Lys  
665 670 675  
Leu Gln Asn Arg Glu Asp Ser Ser Glu Gly Ile Arg Lys Lys Leu  
680 685 690  
Val Glu Ala Glu Glu Arg Arg His Ser Leu Glu Asn Lys Val Lys  
695 700 705

Arg Leu Glu Thr Met Glu Arg Arg Glu Asn Arg Leu Lys Asp Asp  
 710 715 720  
 Ile Gln Thr Lys Ser Gln Gln Ile Gln Gln Met Ala Asp Lys Ile  
 725 730 735  
 Leu Glu Leu Glu Glu Lys His Arg Glu Ala Gln Val Ser Ala Gln  
 740 745 750  
 His Leu Glu Val His Leu Lys Gln Lys Glu Gln His Tyr Glu Glu  
 755 760 765  
 Lys Ile Lys Val Leu Asp Asn Gln Ile Lys Lys Asp Leu Ala Asp  
 770 775 780  
 Lys Glu Thr Leu Glu Asn Met Met Gln Arg His Glu Glu Ala  
 785 790 795  
 His Glu Lys Gly Lys Ile Leu Ser Glu Gln Lys Ala Met Ile Asn  
 800 805 810  
 Ala Met Asp Ser Lys Ile Arg Ser Leu Glu Gln Arg Ile Val Glu  
 815 820 825  
 Leu Ser Glu Ala Asn Lys Leu Ala Ala Asn Ser Ser Leu Phe Thr  
 830 835 840  
 Gln Arg Asn Met Lys Ala Gln Glu Glu Met Ile Ser Glu Leu Arg  
 845 850 855  
 Gln Gln Lys Phe Tyr Leu Glu Thr Gln Ala Gly Lys Leu Glu Ala  
 860 865 870  
 Gln Asn Arg Lys Leu Glu Glu Gln Leu Glu Lys Ile Ser His Gln  
 875 880 885  
 Asp His Ser Asp Lys Asn Arg Leu Leu Glu Leu Glu Thr Arg Leu  
 890 895 900  
 Arg Glu Val Ser Leu Glu His Glu Glu Gln Lys Leu Glu Leu Lys  
 905 910 915  
 Arg Gln Leu Thr Glu Leu Gln Leu Ser Leu Gln Glu Arg Glu Ser  
 920 925 930  
 Gln Leu Thr Ala Leu Gln Ala Ala Arg Ala Ala Leu Glu Ser Gln  
 935 940 945  
 Leu Arg Gln Ala Lys Thr Glu Leu Glu Glu Thr Thr Ala Glu Ala  
 950 955 960  
 Glu Glu Glu Ile Gln Ala Leu Thr Ala His Arg Asp Glu Ile Gln  
 965 970 975  
 Arg Lys Phe Asp Ala Leu Arg Asn Ser Cys Thr Val Ile Thr Asp  
 980 985 990  
 Leu Glu Glu Gln Leu Asn Gln Leu Thr Glu Asp Asn Ala Glu Leu  
 995 1000 1005  
 Asn Asn Gln Asn Phe Tyr Leu Ser Lys Gln Leu Asp Glu Ala Ser  
 1010 1015 1020  
 Gly Ala Asn Asp Glu Ile Val Gln Leu Arg Ser Glu Val Asp His  
 1025 1030 1035  
 Leu Arg Arg Glu Ile Thr Glu Arg Glu Met Gln Leu Thr Ser Gln  
 1040 1045 1050  
 Lys Gln Thr Met Glu Ala Leu Lys Thr Thr Cys Thr Met Leu Glu  
 1055 1060 1065  
 Glu Gln Val Met Asp Leu Glu Ala Leu Asn Asp Glu Leu Leu Glu  
 1070 1075 1080  
 Lys Glu Arg Gln Trp Glu Ala Trp Arg Ser Val Leu Gly Asp Glu  
 1085 1090 1095  
 Lys Ser Gln Phe Glu Cys Arg Val Arg Glu Leu Gln Arg Met Leu  
 1100 1105 1110  
 Asp Thr Glu Lys Gln Ser Arg Ala Arg Ala Asp Gln Arg Ile Thr  
 1115 1120 1125

Glu Ser Arg Gln Val Val Glu Leu Ala Val Lys Glu His Lys Ala  
 1130 1135 1140  
 Glu Ile Leu Ala Leu Gln Gln Ala Leu Lys Glu Gln Lys Leu Lys  
 1145 1150 1155  
 Ala Glu Ser Leu Ser Asp Lys Leu Asn Asp Leu Glu Lys Lys His  
 1160 1165 1170  
 Ala Met Leu Glu Met Asn Ala Arg Ser Leu Gln Gln Lys Leu Glu  
 1175 1180 1185  
 Thr Glu Arg Glu Leu Lys Gln Arg Leu Leu Glu Glu Gln Ala Lys  
 1190 1195 1200  
 Leu Gln Gln Gln Met Asp Leu Gln Lys Asn His Ile Phe Arg Leu  
 1205 1210 1215  
 Thr Gln Gly Leu Gln Glu Ala Leu Asp Arg Ala Asp Leu Leu Lys  
 1220 1225 1230  
 Thr Glu Arg Ser Asp Leu Glu Tyr Gln Leu Glu Asn Ile Gln Val  
 1235 1240 1245  
 Leu Tyr Ser His Glu Lys Val Lys Met Glu Gly Thr Ile Ser Gln  
 1250 1255 1260  
 Gln Thr Lys Leu Ile Asp Phe Leu Gln Ala Lys Met Asp Gln Pro  
 1265 1270 1275  
 Ala Lys Lys Lys Val Pro Leu Gln Tyr Asn Glu Leu Lys Leu  
 1280 1285 1290  
 Ala Leu Glu Lys Glu Lys Ala Arg Cys Ala Glu Leu Glu Glu Ala  
 1295 1300 1305  
 Leu Gln Lys Thr Arg Ile Glu Leu Arg Ser Ala Arg Glu Glu Ala  
 1310 1315 1320  
 Ala His Arg Lys Ala Thr Asp His Pro His Pro Ser Thr Pro Ala  
 1325 1330 1335  
 Thr Ala Arg Gln Gln Ile Ala Met Ser Ala Ile Val Arg Ser Pro  
 1340 1345 1350  
 Glu His Gln Pro Ser Ala Met Ser Leu Leu Ala Pro Pro Ser Ser  
 1355 1360 1365  
 Arg Arg Lys Glu Ser Ser Thr Pro Glu Glu Phe Ser Arg Arg Leu  
 1370 1375 1380  
 Lys Glu Arg Met His His Asn Ile Pro His Arg Phe Asn Val Gly  
 1385 1390 1395  
 Leu Asn Met Arg Ala Thr Lys Cys Ala Val Cys Leu Asp Thr Val  
 1400 1405 1410  
 His Phe Gly Arg Gln Ala Ser Lys Cys Leu Glu Cys Gln Val Met  
 1415 1420 1425  
 Cys His Pro Lys Cys Ser Thr Cys Leu Pro Ala Thr Cys Gly Leu  
 1430 1435 1440  
 Pro Ala Glu Tyr Ala Thr His Phe Thr Glu Ala Phe Cys Arg Asp  
 1445 1450 1455  
 Lys Met Asn Ser Pro Gly Leu Gln Thr Lys Glu Pro Ser Ser Ser  
 1460 1465 1470  
 Leu His Leu Glu Gly Trp Met Lys Val Pro Arg Asn Asn Lys Arg  
 1475 1480 1485  
 Gly Gln Gln Gly Trp Asp Arg Lys Tyr Ile Val Leu Glu Gly Ser  
 1490 1495 1500  
 Lys Val Leu Ile Tyr Asp Asn Glu Ala Arg Glu Ala Gly Gln Arg  
 1505 1510 1515  
 Pro Val Glu Glu Phe Glu Leu Cys Leu Pro Asp Gly Asp Val Ser  
 1520 1525 1530  
 Ile His Gly Ala Val Gly Ala Ser Glu Leu Ala Asn Thr Ala Lys  
 1535 1540 1545

Ala Asp Val Pro Tyr Ile Leu Lys Met Glu Ser His Pro His Thr  
 1550 1555 1560  
 Thr Cys Trp Pro Gly Arg Thr Leu Tyr Leu Leu Ala Pro Ser Phe  
 1565 1570 1575  
 Pro Asp Lys Gln Arg Trp Val Thr Ala Leu Glu Ser Val Val Ala  
 1580 1585 1590  
 Gly Gly Arg Val Ser Arg Glu Lys Ala Glu Ala Asp Ala Lys Leu  
 1595 1600 1605  
 Leu Gly Asn Ser Leu Leu Lys Leu Glu Gly Asp Asp Arg Leu Asp  
 1610 1615 1620  
 Met Asn Cys Thr Leu Pro Phe Ser Asp Gln Val Val Leu Val Gly  
 1625 1630 1635  
 Thr Glu Glu Gly Leu Tyr Ala Leu Asn Val Leu Lys Asn Ser Leu  
 1640 1645 1650  
 Thr His Val Pro Gly Ile Gly Ala Val Phe Gln Ile Tyr Ile Ile  
 1655 1660 1665  
 Lys Asp Leu Glu Lys Leu Leu Met Ile Ala Gly Glu Glu Arg Ala  
 1670 1675 1680  
 Leu Cys Leu Val Asp Val Lys Lys Val Lys Gln Ser Leu Ala Gln  
 1685 1690 1695  
 Ser His Leu Pro Ala Gln Pro Asp Ile Ser Pro Asn Ile Phe Glu  
 1700 1705 1710  
 Ala Val Lys Gly Cys His Leu Phe Gly Ala Gly Lys Ile Glu Asn  
 1715 1720 1725  
 Gly Leu Cys Ile Cys Ala Ala Met Pro Ser Lys Val Val Ile Leu  
 1730 1735 1740  
 Arg Tyr Asn Glu Asn Leu Ser Lys Tyr Cys Ile Arg Lys Glu Ile  
 1745 1750 1755  
 Glu Thr Ser Glu Pro Cys Ser Cys Ile His Phe Thr Asn Tyr Ser  
 1760 1765 1770  
 Ile Leu Ile Gly Thr Asn Lys Phe Tyr Glu Ile Asp Met Lys Gln  
 1775 1780 1785  
 Tyr Thr Leu Glu Glu Phe Leu Asp Lys Asn Asp His Ser Leu Ala  
 1790 1795 1800  
 Pro Ala Val Phe Ala Ala Ser Ser Asn Ser Phe Pro Val Ser Ile  
 1805 1810 1815  
 Val Gln Val Asn Ser Ala Gly Gln Arg Glu Glu Tyr Leu Leu Cys  
 1820 1825 1830  
 Phe His Glu Phe Gly Val Phe Val Asp Ser Tyr Gly Arg Arg Ser  
 1835 1840 1845  
 Arg Thr Asp Asp Leu Lys Trp Ser Arg Leu Pro Leu Ala Phe Ala  
 1850 1855 1860  
 Tyr Arg Glu Pro Tyr Leu Phe Val Thr His Phe Asn Ser Leu Glu  
 1865 1870 1875  
 Val Ile Glu Ile Gln Ala Arg Ser Ser Ala Gly Thr Pro Ala Arg  
 1880 1885 1890  
 Ala Tyr Leu Asp Ile Pro Asn Pro Arg Tyr Leu Gly Pro Ala Ile  
 1895 1900 1905  
 Ser Ser Gly Ala Ile Tyr Leu Ala Ser Ser Tyr Gln Asp Lys Leu  
 1910 1915 1920  
 Arg Val Ile Cys Cys Lys Gly Asn Leu Val Lys Glu Ser Gly Thr  
 1925 1930 1935  
 Glu His His Arg Gly Pro Ser Thr Ser Arg Ser Ser Pro Asn Lys  
 1940 1945 1950  
 Arg Gly Pro Pro Thr Tyr Asn Glu His Ile Thr Lys Arg Val Ala  
 1955 1960 1965

Ser Ser Pro Ala Pro Pro Glu Gly Pro Ser His Pro Arg Glu Pro  
 1970 1975 1980  
 Ser Thr Pro His Arg Tyr Arg Glu Gly Arg Thr Glu Leu Arg Arg  
 1985 1990 1995  
 Asp Lys Ser Pro Gly Arg Pro Leu Glu Arg Glu Lys Ser Pro Gly  
 2000 2005 2010  
 Arg Met Leu Ser Thr Arg Arg Glu Arg Ser Pro Gly Arg Leu Phe  
 2015 2020 2025  
 Glu Asp Ser Ser Arg Gly Arg Leu Pro Ala Gly Ala Val Arg Thr  
 2030 2035 2040  
 Pro Leu Ser Gln Val Asn Lys Val Trp Asp Gln Ser Ser Val  
 2045 2050

&lt;210&gt; 22

&lt;211&gt; 1665

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7638121CD1

&lt;400&gt; 22

Met Gly Cys Cys Arg Leu Gly Cys Gly Gly Cys Ser Val Ala His  
 1 5 10 15  
 Ser Val Ser Gln Gly Leu Thr Asn His Pro Ser Met Val Gly Cys  
 20 25 30  
 Gly Trp His Pro Gly Leu Cys Gly Trp Gly Gly Gly Leu His Ser  
 35 40 45  
 Ser Leu Pro Ala Leu Pro Gly Pro Pro Ser Met Gln Val Thr Ile  
 50 55 60  
 Glu Asp Val Gln Ala Gln Thr Gly Gly Thr Ala Gln Phe Glu Ala  
 65 70 75  
 Ile Ile Glu Gly Asp Pro Gln Pro Ser Val Thr Trp Tyr Lys Asp  
 80 85 90  
 Ser Val Gln Leu Val Asp Ser Thr Arg Leu Ser Gln Gln Glu  
 95 100 105  
 Gly Thr Thr Tyr Ser Leu Val Leu Arg His Met Ala Ser Lys Asp  
 110 115 120  
 Ala Gly Val Tyr Thr Cys Leu Ala Gln Asn Thr Gly Gly Gln Val  
 125 130 135  
 Leu Cys Lys Ala Glu Leu Leu Val Leu Gly Gly Asp Asn Glu Pro  
 140 145 150  
 Asp Ser Glu Lys Gln Ser His Arg Arg Lys Leu His Ser Phe Tyr  
 155 160 165  
 Glu Val Lys Glu Glu Ile Gly Arg Gly Val Phe Gly Phe Val Lys  
 170 175 180  
 Arg Val Gln His Lys Gly Asn Lys Ile Leu Cys Ala Ala Lys Phe  
 185 190 195  
 Ile Pro Leu Arg Ser Arg Thr Arg Ala Gln Ala Tyr Arg Glu Arg  
 200 205 210  
 Asp Ile Leu Ala Ala Leu Ser His Pro Leu Val Thr Gly Leu Leu  
 215 220 225  
 Asp Gln Phe Glu Thr Arg Lys Thr Leu Ile Leu Ile Leu Glu Leu  
 230 235 240  
 Cys Ser Ser Glu Glu Leu Leu Asp Arg Leu Tyr Arg Lys Gly Val

245	250	255
Val Thr Glu Ala Glu Val Lys Val Tyr Ile Gln Gln Leu Val Glu		
260	265	270
Gly Leu His Tyr Leu His Ser His Gly Val Leu His Leu Asp Ile		
275	280	285
Lys Pro Ser Asn Ile Leu Met Val His Pro Ala Arg Glu Asp Ile		
290	295	300
Lys Ile Cys Asp Phe Gly Phe Ala Gln Asn Ile Thr Pro Ala Glu		
305	310	315
Leu Gln Phe Ser Gln Tyr Gly Ser Pro Glu Phe Val Ser Pro Glu		
320	325	330
Ile Ile Gln Gln Asn Pro Val Ser Glu Ala Ser Asp Ile Trp Ala		
335	340	345
Met Gly Val Ile Ser Tyr Leu Ser Leu Thr Cys Ser Ser Pro Phe		
350	355	360
Ala Gly Glu Ser Asp Arg Ala Thr Leu Leu Asn Val Leu Glu Gly		
365	370	375
Arg Val Ser Trp Ser Ser Pro Met Ala Ala His Leu Ser Glu Asp		
380	385	390
Ala Lys Asp Phe Ile Lys Ala Thr Leu Gln Arg Ala Pro Gln Ala		
395	400	405
Arg Pro Ser Ala Ala Gln Cys Leu Ser His Pro Trp Phe Leu Lys		
410	415	420
Ser Met Pro Ala Glu Glu Ala His Phe Ile Asn Thr Lys Gln Leu		
425	430	435
Lys Phe Leu Leu Ala Arg Ser Arg Trp Gln Arg Ser Leu Met Ser		
440	445	450
Tyr Lys Ser Ile Leu Val Met Arg Ser Ile Pro Glu Leu Leu Arg		
455	460	465
Gly Pro Pro Asp Ser Pro Ser Leu Gly Val Ala Arg His Leu Cys		
470	475	480
Arg Asp Thr Gly Gly Ser Ser Ser Ser Ser Ser Ser Asp Asn		
485	490	495
Glu Leu Ala Pro Phe Ala Arg Ala Lys Ser Leu Pro Pro Ser Pro		
500	505	510
Val Thr His Ser Pro Leu Leu His Pro Arg Gly Phe Leu Arg Pro		
515	520	525
Ser Ala Ser Leu Pro Glu Glu Ala Glu Ala Ser Glu Arg Ser Thr		
530	535	540
Glu Ala Pro Ala Pro Pro Ala Ser Pro Glu Gly Ala Gly Pro Pro		
545	550	555
Ala Ala Gln Gly Cys Val Pro Arg His Ser Val Ile Arg Ser Leu		
560	565	570
Phe Tyr His Gln Ala Gly Glu Ser Pro Glu His Gly Ala Leu Ala		
575	580	585
Pro Gly Ser Arg Arg His Pro Ala Arg Arg Arg His Leu Leu Lys		
590	595	600
Gly Gly Tyr Ile Ala Gly Ala Leu Pro Gly Leu Arg Glu Pro Leu		
605	610	615
Met Glu His Arg Val Leu Glu Glu Ala Ala Arg Glu Glu Gln		
620	625	630
Ala Thr Leu Leu Ala Lys Ala Pro Ser Phe Glu Thr Ala Leu Arg		
635	640	645
Leu Pro Ala Ser Gly Thr His Leu Ala Pro Gly His Ser His Ser		
650	655	660
Leu Glu His Asp Ser Pro Ser Thr Pro Arg Pro Ser Ser Glu Ala		

665	670	675
Cys Gly Glu Ala Gln Arg Leu Pro Ser Ala Pro Ser Gly Gly Ala		
680	685	690
Pro Ile Arg Asp Met Gly His Pro Gln Gly Ser Lys Gln Leu Pro		
695	700	705
Ser Thr Gly Gly His Pro Gly Thr Ala Gln Pro Glu Arg Pro Ser		
710	715	720
Pro Asp Ser Pro Trp Gly Gln Pro Ala Pro Phe Cys His Pro Lys		
725	730	735
Gln Gly Ser Ala Pro Gln Glu Gly Cys Ser Pro His Pro Ala Val		
740	745	750
Ala Pro Cys Pro Pro Gly Ser Phe Pro Pro Gly Ser Cys Lys Glu		
755	760	765
Ala Pro Leu Val Pro Ser Ser Pro Phe Leu Gly Gln Pro Gln Ala		
770	775	780
Pro Leu Ala Pro Ala Lys Ala Ser Pro Pro Leu Asp Ser Lys Met		
785	790	795
Gly Pro Gly Asp Ile Ser Leu Pro Gly Arg Pro Lys Pro Gly Pro		
800	805	810
Cys Ser Ser Pro Gly Ser Ala Ser Gln Ala Ser Ser Ser Gln Val		
815	820	825
Ser Ser Leu Arg Val Gly Ser Ser Gln Val Gly Thr Glu Pro Gly		
830	835	840
Pro Ser Leu Asp Ala Glu Gly Trp Thr Gln Glu Ala Glu Asp Leu		
845	850	855
Ser Asp Ser Thr Pro Thr Leu Gln Arg Pro Gln Glu Gln Val Thr		
860	865	870
Met Arg Lys Phe Ser Leu Gly Gly Arg Gly Tyr Ala Gly Val		
875	880	885
Ala Gly Tyr Gly Thr Phe Ala Phe Gly Gly Asp Ala Gly Gly Met		
890	895	900
Leu Gly Gln Gly Pro Met Trp Ala Arg Ile Ala Trp Ala Val Ser		
905	910	915
Gln Ser Glu Glu Glu Glu Gln Glu Ala Arg Ala Glu Ser Gln		
920	925	930
Ser Glu Glu Gln Glu Ala Arg Ala Glu Ser Pro Leu Pro Gln		
935	940	945
Val Ser Ala Arg Pro Val Pro Glu Val Gly Arg Ala Pro Thr Arg		
950	955	960
Ser Ser Pro Glu Pro Thr Pro Trp Glu Asp Ile Gly Gln Val Ser		
965	970	975
Leu Val Gln Ile Arg Asp Leu Ser Gly Asp Ala Glu Ala Ala Asp		
980	985	990
Thr Ile Ser Leu Asp Ile Ser Glu Val Asp Pro Ala Tyr Leu Asn		
995	1000	1005
Leu Ser Asp Leu Tyr Asp Ile Lys Tyr Leu Pro Phe Glu Phe Met		
1010	1015	1020
Ile Phe Arg Lys Val Pro Lys Ser Ala Gln Pro Glu Pro Pro Ser		
1025	1030	1035
Pro Met Ala Glu Glu Glu Leu Ala Glu Phe Pro Glu Pro Thr Trp		
1040	1045	1050
Pro Trp Pro Gly Glu Leu Gly Pro His Ala Gly Leu Glu Ile Thr		
1055	1060	1065
Glu Glu Ser Glu Asp Val Asp Ala Leu Leu Ala Glu Ala Ala Val		
1070	1075	1080
Gly Arg Lys Arg Lys Trp Ser Ser Pro Ser Arg Ser Leu Phe His		

1085	1090	1095
Phe Pro Gly Arg His Leu Pro Leu Asp Glu Pro Ala Glu Leu Gly		
1100	1105	1110
Leu Arg Glu Arg Val Lys Ala Ser Val Glu His Ile Ser Arg Ile		
1115	1120	1125
Leu Lys Gly Arg Pro Glu Gly Leu Glu Lys Glu Gly Pro Pro Arg		
1130	1135	1140
Lys Lys Pro Gly Leu Ala Ser Phe Arg Leu Ser Gly Leu Lys Ser		
1145	1150	1155
Trp Asp Arg Ala Pro Thr Phe Leu Arg Glu Leu Ser Asp Glu Thr		
1160	1165	1170
Val Val Leu Gly Gln Ser Val Thr Leu Ala Cys Gln Val Ser Ala		
1175	1180	1185
Gln Pro Ala Ala Gln Ala Thr Trp Ser Lys Asp Gly Ala Pro Leu		
1190	1195	1200
Glu Ser Ser Ser Arg Val Leu Ile Ser Ala Thr Leu Lys Asn Phe		
1205	1210	1215
Gln Leu Leu Thr Ile Leu Val Val Ala Glu Asp Leu Gly Val		
1220	1225	1230
Tyr Thr Cys Ser Val Ser Asn Ala Leu Gly Thr Val Thr Thr Thr		
1235	1240	1245
Gly Val Leu Arg Lys Ala Glu Arg Pro Ser Ser Ser Pro Cys Pro		
1250	1255	1260
Asp Ile Gly Glu Val Tyr Ala Asp Gly Val Leu Leu Val Trp Lys		
1265	1270	1275
Pro Val Glu Ser Tyr Gly Pro Val Thr Tyr Ile Val Gln Cys Ser		
1280	1285	1290
Leu Glu Gly Gly Ser Trp Thr Thr Leu Ala Ser Asp Ile Phe Asp		
1295	1300	1305
Cys Cys Tyr Leu Thr Ser Lys Leu Ser Arg Gly Gly Thr Tyr Thr		
1310	1315	1320
Phe Arg Thr Ala Cys Val Ser Lys Ala Gly Met Gly Pro Tyr Ser		
1325	1330	1335
Ser Pro Ser Glu Gln Val Leu Leu Gly Gly Pro Ser His Leu Ala		
1340	1345	1350
Ser Glu Glu Glu Ser Gln Gly Arg Ser Ala Gln Pro Leu Pro Ser		
1355	1360	1365
Thr Lys Thr Phe Ala Phe Gln Thr Gln Ile Gln Arg Gly Arg Phe		
1370	1375	1380
Ser Val Val Arg Gln Cys Trp Glu Lys Ala Ser Gly Arg Ala Leu		
1385	1390	1395
Ala Ala Lys Ile Ile Pro Tyr His Pro Lys Asp Lys Thr Ala Val		
1400	1405	1410
Leu Arg Glu Tyr Glu Ala Leu Lys Gly Leu Arg His Pro His Leu		
1415	1420	1425
Ala Gln Leu His Ala Ala Tyr Leu Ser Pro Arg His Leu Val Leu		
1430	1435	1440
Ile Leu Glu Leu Cys Ser Gly Pro Glu Leu Leu Pro Cys Leu Ala		
1445	1450	1455
Glu Arg Ala Ser Tyr Ser Glu Ser Glu Val Lys Asp Tyr Leu Trp		
1460	1465	1470
Gln Met Leu Ser Ala Thr Gln Tyr Leu His Asn Gln His Ile Leu		
1475	1480	1485
His Leu Asp Leu Arg Ser Glu Asn Met Ile Ile Thr Glu Tyr Asn		
1490	1495	1500
Leu Leu Lys Val Val Asp Leu Gly Asn Ala Gln Ser Leu Ser Gln		

1505	1510	1515
Glu Lys Val Leu Pro Ser Asp Lys Phe Lys Asp Tyr Leu Glu Thr		
1520	1525	1530
Met Ala Pro Glu Leu Leu Glu Gly Gln Gly Ala Val Pro Gln Thr		
1535	1540	1545
Asp Ile Trp Ala Ile Gly Val Thr Ala Phe Ile Met Leu Ser Ala		
1550	1555	1560
Glu Tyr Pro Val Ser Ser Glu Gly Ala Arg Asp Leu Gln Arg Gly		
1565	1570	1575
Leu Arg Lys Gly Leu Val Arg Leu Ser Arg Cys Tyr Ala Gly Leu		
1580	1585	1590
Ser Gly Gly Ala Val Ala Phe Leu Arg Ser Thr Leu Cys Ala Gln		
1595	1600	1605
Pro Trp Gly Arg Pro Cys Ala Ser Ser Cys Leu Gln Cys Pro Trp		
1610	1615	1620
Leu Thr Glu Glu Gly Pro Ala Cys Ser Arg Pro Ala Pro Val Thr		
1625	1630	1635
Phe Pro Thr Ala Arg Leu Arg Val Phe Val Arg Asn Arg Glu Lys		
1640	1645	1650
Arg Arg Ala Leu Leu Tyr Lys Arg His Asn Leu Ala Gln Val Arg		
1655	1660	1665

<210> 23

<211> 1014

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7482896CB1

<400> 23

atgacaaaca acagcggctc caaagccaa ctcgttgg gaggaaata caaaactggtg 60  
 cggaaagatcg ggtctggctc ctttggagac gtttatctgg gcatcaccac caccacggc 120  
 gaggacgtag cagtgaagct ggaatctcg aaggtcaagc acccccagtt gctgtatgag 180  
 agcaaactct acacgattct tcaagggtggg gttggcatcc cccacatgca ctgttatgg 240  
 cagaaaaaaag acaacaatgt gctagtcatg gaccttctgg gaccacgctc cgaagacctc 300  
 tttaatttct gttcaagaag gttcaccatg aaaactgtac ttatgttagc cgaccagatg 360  
 atcagcagaa ttgaatacgt gcatacaaag aattttctac accgagacat taaaccagat 420  
 aacttcctga tgggtactgg gctgtactgt aataagttgt tccttattga ttttggtttg 480  
 gccaaaaagt acagagacaa caggaccagg caacacatac cgtacagaga agataaacac 540  
 ctcattggca ctgtccgata tgccagcatc aatgcacatc ttggatttga gcagagccgc 600  
 cgagatgaca tggaaatcctt aggctacgtt ttcatgtatt ttaatagaac cagcctgccc 660  
 tggcaaggac taagggctat gacaaaaaaa caaaaatatg aaaagattag tgagaagaag 720  
 atgtccaccc ctgttgaagt ttatgtaaag gggtttcctg cagaattcgc catgtacttg 780  
 aactactgtc gtgggctgca ctttgaggaa gtcccagatt acatgtatct gaggcagcta 840  
 ttccgcattc tttttaggac cctgaaccac caatatgact acacatttga ttggacgtatg 900  
 ttaaagcaga aagcagcaca gcaggcagcc tcttccagtg ggcagggtca gcaggcccaa 960  
 acccagacag gcaagcaaac tgaaaaaaac aagaataatg tgaaaagataa ctaa 1014

<210> 24

<211> 1530

<212> DNA

<213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7483046CB1

<400> 24

cggcctgaca ggccggcatg cggcgccca gactgtagcc gagcagcgag gctccggccg 60  
 cagccatgga gcggcgctg cgccgcgtgg agcagctggc gcggggcgag gccggcgct 120  
 gccccgggct cgacggcctc ctagatctgc tgcgtggcgct gcaccacgag cttagcagcg 180  
 gccccctacg gcgggagcgc agcgtggcgc agttccttag ctgggcccgc cccttcgtat 240  
 caaaggtaaa agaactgcgt ctgcagagag atgactttga gatcttgaag gtgatcggcc 300  
 gaggagcctt tggggaggc accgtggtga ggcagaggga cactggcag attttgcca 360  
 tgaaaatgtc gcacaagtgg gagatgtcga agagggctga gacagcctgt ttccgggagg 420  
 agcgggatgt gctcgtgaaa ggggacagcc gttgggtgac cactctgcac tatgccttcc 480  
 aagacgagga gtacctgtac cttgtatgg actactatgc tgggtgggac ctcctgacgc 540  
 tgctgagccg cttcgaggac cgtctccgc cccagctggc ccagttctac ctggctgaga 600  
 tggtgctggc catccactcg ctgcaccaggc tgggttatgt ccacaggat gtcaagccag 660  
 acaacgtcct gctggatgtg aacgggacaca ttccctggc tgacttcggc tcctgcctgc 720  
 gtctcaacac caacggcatg gtggattcat cagttggcagt agggacgccc gactatatct 780  
 cccctgagat cctgcaggcc atggaggagg gcaaggccca ctacggccca cagtgtgact 840  
 ggtggtcgt tggagtctgc gcctatgagc tgctctttgg ggagacgccc ttctatgtc 900  
 agtccttggt gaaaacctac ggcaagatca tgaaccacga ggaccacctg cagttcccc 960  
 cggacgtgcc tgacgtgcca gccagcgcac aagacctgtat ccggccagctg ctgtgtcgcc 1020  
 aggaagagcg gctaggccgt ggtgggtctgg atgacttccg gaaccatctt ttcttcgaag 1080  
 gcgtggactg ggagcggctg gcgagcagca cggccccccta tattccttag ctgcggggc 1140  
 ccatggacac ctccaacttt gatgtggatg acgacaccct caaccatcca gggaccctgc 1200  
 caccggccctc ccacggggcc ttctccggcc atcacctgccc attcgtggc ttcacctaca 1260  
 cctcaggcag tcacagtctt gagagcagct ctgaggcttgg ggctgcctg gagcggaaagc 1320  
 tccagtgtct ggagcaggag aagggtggagc tgagcaggaa gcaccaagag gcctgcacg 1380  
 ccccccacaga ccatcggggag ctggagcagc tacggaaagg agtgcagact ctgcgggaca 1440  
 ggctgcccagg tatcccttcc gcccacccccc accctctcct tgagttctg tgaattaaaa 1500  
 tatttgc当地 tccaaaaaaaaaaaaaaaagg 1530

<210> 25  
 <211> 3150  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 71636374CB1

<400> 25

atggccttat agaaaaatt gatttataaa aagtggtaca ggtttcata gataaccatg 60  
 acaacatccc atatgaatgg gcatgttaca gaggaaatcg acagcgaagt aaaaaatgtt 120  
 gatcttgcac caccagagga acatcagaag caccgagaga tggctgttga ctgcctggc 180  
 gatttggca ccaggatgtat gccaatacgt cgaagtgcac agttggagcg tattcggcaa 240  
 caacaggagg acatgaggcg taggagagag gaagaaggaa aaaagcaaga acttgacctt 300  
 aattcttcca tgagacttaa gaaactagcc caaattcctc caaagaccgg aatagataac 360  
 cctatgtttt atacagagga aggaattgtc ttagaaagtc ctcattatgc tgtaaaata 420  
 ttagaaatag aagacttgc ttcttcactt aaacatatacc aacatacttt ggtagattct 480  
 cagagccagg aggatatttc actgctttca caacttgcggc aaaataagga ttccagaat 540  
 gcatttaaga tacacaatgc catcacagta cacatgaaca aggccagttcc tccatttcct 600  
 cttatcttcca acgcacaaga tcttgctcaa gagttacaaa ctgtttgaa gccagttcat 660  
 cataaggaag gacaagaact aactgctttc ctgaataactc cacatattca ggcactttta 720  
 ctggcccacg ataaggtgc tgagcaggaa atgcagctg agcccatc agatgagaga 780  
 gtttatgaaa gtattggcca gtatggagga gaaactgtaa aaatagttcg tatagaaaaag 840

gctcgtgata ttccgttggg tgctacagtt cgtaatgaaa tggactctgt catcatttgc 900  
 cggatagataa aagggggtgc tgcagagaaa agtggctgt tgcatgaagg agatgaagtt 960  
 ctagagatta atggcattga aattcggggg aaagatgtca atgaggttt tgacttgg 1020  
 tctgatatgc atggtaactt gactttgtc ctgattccca gtcaacagat caagccgc 1080  
 cctgccaagg aaacagtaat ccatgtaaaa gtcattttg actatgaccc ctcagatgac 1140  
 ccttatgttc catgtcgaga gttaggtctg tctttcaaa aagggtatat acttcatgtg 1200  
 atcagtcaag aagatccaaa ctggtggcag gcctacaggg aaggggacga agataatcaa 1260  
 cctctagccg ggcttggcc agggaaaagc tttagcagc aaagggaaagc catgaaacaa 1320  
 accatagaag aagataagga gccagaaaaa tcaggaaaac tgggtgtgc aaagaagaat 1380  
 aaaaagaaga gaaaaaagggt ttatataat gccaataaaa atgatgatta tgacaacgag 1440  
 gagatcttaa cctatgagga aatgtcaattt tattatcagc cagaaatag gaagagac 1500  
 atcatcttgc ttggtccaca gaactgtggc cagaatgaat tgcgtcagag gctcatgaaac 1560  
 aaagaaaaagg accgcttgc atctgcattt cctcatacaa cccggagtag gcgagaccaa 1620  
 gaagtagccg gtagagatta ccactttgtt tcgcggcaag cattcgaggc agacatagca 1680  
 gctggaaagt tcattgagca tggtaattt gagaagaatt tggtaattt tagcatagat 1740  
 tctgtacggc aagtgtcaat ctctggcaaa atatgtctt taagtcttcg tacacagtca 1800  
 ttgaagactc tccggattt agatttggaaa ccatatatta tcttcatttc acccccttca 1860  
 caagaaagac ttccggcatt attggccaaa gaaggcaaga atccaaagcc tgaagagtt 1920  
 agagaaatca ttgagaagac aagagagatg gaggcagaaca atggccacta ctttgatacg 1980  
 gcaatttgta attccgatct tgataaaagcc tattatcattt tgcttaggtt aattaacaaa 2040  
 cttgatactg aacccatgt ggtaccatcc acttggctga ggtggaaacccatcc 2100  
 tggcatgt tggacttgat ctggcaaaaaa ctggcaatag gaggactgccc cgacactgca 2160  
 gcaagattga ggataagatg gaaggcagca gtataagctg tagatctgtt cttagatctc 2220  
 ttgaatttagt gagacgacag ttccctttagg cagtttgc atggcatctt ttattctctt 2280  
 tacatggctt tagcggttct tgcctcattt tggattcta aatggaaatctt tcaacagag 2340  
 cattccattt tgcctgttta aaaccttttgc tttcaccta aaccctttct gcttagttt 2400  
 atctctgtca aaaacttgc tacacaagcg tccatgtctc acacaaatattt tgatgtgatt 2460  
 attcttaagt gttaaatcat taacacttta atgacttcat tggaaatattt gagcagaggg 2520  
 actgtgttc tatgcactgg gcaaggcagt atttgcttag gaaactaattt tagtcatcag 2580  
 agataacttc ctaaaaaagga aaaataaaaaa aaaaaatggt gccactttgg gttgaagcta 2640  
 ctttgttagg cttagattca tttatatgtc ttttgattttt taaaaaaaca aaaaacattt 2700  
 cattagaagc accagttttt ttgctcagac tttgtggatc agactctaca ctcaacacac 2760  
 tctaatttac taaaaggat taaaatattt ctgatctttt taaaattatg atttcctgaa 2820  
 tttttttctt aagtcgtctc aactgatttt ctcacttagc ttcccttccc tcattcagcat 2880  
 agtataatag aatgtatgtt acatttttt gatggcagg tggatattt aatctgtatt 2940  
 gactaaaaaa gtttcttcct catgatgcta atagttttt gtatacatgg gaggatagca 3000  
 catttgacag ttttgcatt tttatgtatg agcacaatgtt cctatgactg tgctacgtat 3060  
 atataagttaa taaaactggaa ttctgttgcattt gatataatgtt gctgtactgtt atattaat 3120  
 ttaatagatc aacaaatggt cattgaaaaac 3150

<210> 26  
 <211> 2901  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7480597CB1

<400> 26  
 atggcggaaag gcaagggaaagg gcaaggccca tcttacatgg atggcagcag gcaaagagag 60  
 aatgagaaag atgcaaaagc ggaaacccctt gatgtacca tcagatctt tgatgtttt 120  
 tcactaccat ggaacacagaca gcaaggccca tggaccatt ctctaaaata tttaaatgtcg 180  
 agaatacacag agcggaaagct gcaaggctcc tggctgcctg ccagccgagg gaatctggag 240  
 aaaccattcc tggggccgcg tggccctgtt tctgcccctcg gaatggccctt 300  
 cactcagcac atcctgagaa cagccctctg aagcccaaggc tgcgtgaccgt agtgaagctg 360

ggggggcggc gcccccgaaa gatcaactctg ctccctcaaca ggcgatcagt gcagacgttc 420  
gaggcagctct tagctgacat ctcagaagcc ttgggctctc ccagatggaa gaatgaccgt 480  
gtgagggaaac ttttaacctt caagggcagg gaaatcagga gctgtctctga tttcttcagg 540  
gaaggggatg cttcatagc tatggcaaa gaaccactga cactgaagag cattcaggtg 600  
gctgtagaag aactgtaccc caacaaagcc cggccctga cactggccca gcacagccgt 660  
gcccccttc caaggcttag ggcaggctg ttagcaagg ctctgaaagg agaccaccgc 720  
tgtggggaga ccgagacccc caagagctgc agcgaagttt caggatgcaa ggcagccatg 780  
aggcaccagg ggaagatccc cgaggagctt tcactagatg acagagcgag gaccagaag 840  
aagtggggaa gggggaaatg ggagccagaa cccagtagca agccccccag ggaagccact 900  
ctggaaagaga ggcacgcaag gggagagaag cattttttttggg tggagattga aaagacctcg 960  
ggtgaaattt tcagatgcga gaagtgcag agagagaggg agctccagca ggcctggag 1020  
cgtgagaggc tttctctggg gaccagttag ctggatatgg ggaagggccc aatgtatgtat 1080  
gtggagaagc tggtgaggac cagaagctgc aggaggtctc ccgaggcaaa tcctgcaagt 1140  
ggggagggaa ggttggaaaggg tgacagccac aggagcagcc ccaggaatcc cactcaagag 1200  
ctgaggagac ccagcaagag catggacaag aaagaggaca gaggcccaga ggatcaagaa 1260  
agccatgctc agggagcagc caaggccaag aaggaccttgg tggagttct tcctgtcaca 1320  
gaggaggggc tgagggaggt gaagaaggac accaggccca tgagcaggag caaacatggt 1380  
ggctggctcc tgagagagca ccaggcggc tttgagaagc tccgcaggac ccgaggagaa 1440  
gagaaggagg cagagaagga gaaaaagcca ttttatgtctg gaggcagaag gatgactctc 1500  
agagatgacc aacctgcaaa gctagaaaaag gagcccaaga cgaggccaga agagaacaag 1560  
ccagagcggc ccagcggtcg gaagccacgg cccatggca tcattggcgc caatgtggaa 1620  
aagcattatg agactggccg gtcattttttt gatggaaact ttgctgtcgt gaaggagtgc 1680  
agacaccgcg agaccaggca ggcctatgcg atgaagatca ttgacaagtc cagactcaag 1740  
ggcaaggagg acatggtggat cagttagatc ttgatcatcc aggcctctc tcaccccaac 1800  
atctgttaat tgcattttttt gatggaaaca gacatggaaa tctacgtat cctggagttac 1860  
gtgcaggagg gagacctttt tgacgcccattc atagaaaagtg tgaagtccccccggat 1920  
gctgcctca tgcattttttt gatggaaaca ggcctcgatcc acatgcacga caagagcatt 1980  
gtccaccggg acctcaagcc ggaaaaacctt ttggttcagc gaaatgagga caaatctact 2040  
accttggaaat tggctgatcc tggacttgca aagcatgtgg tgagacccat atttactgt 2100  
tgtgggaccc caacttacgt agctcccgaa attttttctg agaaaaggta tggactggag 2160  
gtggacatgt gggctgctgg cgtgatcttc tatatcttc ttttttttttcccccattc 2220  
cgccagccctg agagggacca ggacgagctc ttttttttttcccccattc 2280  
ttccctcccccc cttaactggta caatatcttc gatgtctgtca aagatctgtt gagccgggtt 2340  
ctgggtgttag accccaaaaaa ggcctacaca gctcatcagg ttcttcagca cccctggatc 2400  
gaaacagctg gcaagaccaa tacagtggaa cgacagaagc aggtgtcccc cagcagcgg 2460  
ggtcacttcc ggagccagca caagagggttt gtggagcagg tatcatagtc accaccttgg 2520  
gaatctgtcc agccccccagt tctgctcaag gacagagaaaa aggtatagaa tttgagagaa 2580  
aaacaatgaa agaggcttct tcacataatt ggtgaatcag agggagagac actgagtata 2640  
ttttaaagca tattttttttt attaagtcaa ttttttttttcccccattc 2700  
gtatattttttaa agccttaat acattttttttgg ggggttaagca ttgtcatcag tgaggaattt 2760  
tggttaataat gatgtgtttt gtttttttttcccccattc 2820  
tggtgcttac cagggtctaa actccccctg tgagattat aaggtgcatt gtggctttc 2880  
tgtgttaataa aatgtggtc c 2901

<210> 27

<211> 1671

<212> DNA

<213> *Homo sapiens*

<220>

<221> misc\_feature

<223> Incyte ID No: 3227248CB1

<400> 27

atgaagctta taaaatggcaa aaagcaaaaca ttcccatggt ttggcatgga catcggtgga 60  
acgctggta aattggtgta tttcgagccg aaggatatta cagccgaaga ggagcaagag 120

gaagtggaga acctgaagag catccgaaag tatttgactt ctaatactgc ttatggaaa 180  
 actgggatcc gagacgtcca cctggaactg aaaaacctga ccatgtgtgg acgcaaagg 240  
 aacctgact tcatccgctt tcccagctgt gctatgcaca ggttcatca gatgggcagc 300  
 gagaagaact tctcttagct tcacaccacc ctctgtgcca caggaggcgg ggcttcaaa 360  
 ttcgaagagg acttcagaat gattgctgac ctgcagctgc ataaaactgga tgaactggac 420  
 tgtctgattc agggcctgct ttatgtcgac tctgttggct tcaacggcaa gccagaatgt 480  
 tactattttg aaaatcccac aaatcctgaa ttgtgtcaaa aaaagccgta ctgccttgat 540  
 aaccatacc ctatgttgct ggttaacatg ggctcaggtg tcagcattct agccgtgtac 600  
 tccaaggaca actataaaaag agttacaggg accagtctg gaggtgaaac attccttaggc 660  
 ctatgttgct tgctgactgg ttgtgagacc tttgaagaag ctctggaaat ggcagctaaa 720  
 ggcgacagca ccaatgttga taaactggtg aaggacattt acggaggaga ctatgaacga 780  
 tttggcccttc aaggatctgc ttagcatca agcttggca acatgatgag taaagaaaag 840  
 cgagattcca tcagcaagga agacctcgcc cgggccacat tggtcaccat caccaacaac 900  
 attggctcca ttgctcggt gtgtgcgtt aatgagaaca tagacagagt tggttttgtt 960  
 gaaaattttc tcagaatcaa tatggctcc atgaagctgc tggcatatgc catggatttt 1020  
 tggtccaaag gacaactgaa agctctgtt ttgaacatg agggtttattt tggagccgtt 1080  
 ggggcactgt tggaaactgtt caaaatgact gatgataagt agagacgagc agtggaggaa 1140  
 acagcctccc aaaaggacag agaactaaaa aattgctgct ggagaagtg aaagtgcgtt 1200  
 tgggacggaa gccaagccat tatggcagat gaacctgctg gatttgtaaa taattttaaa 1260  
 tccttcaga ttagtttta ctcttaggtt tttagctaat gattcaaaac ggggaatata 1320  
 aaaaggtttt tttctgtat actgtatttt tttaaaaaaaa tggtcagcg tggccaaacc 1380  
 taccaattgt atgcattaaac ttgaaaagt tgggtatgt ttaagaagga cctgatatgt 1440  
 aagcgttgtt cattttctt ctggggttta ctgatcgtg tggtatttt aacttcattt 1500  
 agtaattact cttaggagatt ttaccttgac ttatatttt catgacgttt catgatttgc 1560  
 tggggtttc aaatgaaaact acaaacttgg catgtttac tgtgaacact tttgttattt 1620  
 gttttgtacc cttttgtct tgttttctg ttttagttgt cttctgaaaa a 1671

<210> 28

<211> 2577

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 4207273CB1

<400> 28

atgccacaga tagcaaagaa gcaatcaact caccggactc agaaacctaa aaagcaatca 60  
 tttcctgca tctgtaaaaa tccaggaaca cagaagtcat gtgttccctct ctctgttcaa 120  
 ccgacagacg caagactaaa ttaccttagat cttaagtata gtgatatgtt caaagaaatc 180  
 aattcaactg ctaatggacc tggaaatctat gaaatgtttg ggacccctgt ttattgtcat 240  
 gtgcgagaga ctgaaaggga tgaaaacacg tattaccgtg agatatgttgc ggctccatca 300  
 ggcagacgta tcaccaataa atgtcgatct tcacacagtg agaggaagag caatatcaga 360  
 acaagacttt cttagaaaaaa aacacatatg aaatgcccac agacttcatt tggcattaaa 420  
 caagagcaca aagtcttaat ttctaaagaa aagagtccaa aggctgtaca tagcaaccta 480  
 catgacatttggaaaatggta tggatttca gaaccagact ggcagataaa gtttcagga 540  
 aatgagtttca tatcttccaa agatgaaatt catcccatga acttggctca gacacctgag 600  
 cagtccatga aacagaatga attccctcct gtctcagatt tatccattgt tgaagaagtt 660  
 tctatggaag agtctactgg tgatagagac atttctaaaca atcaaataact caccacaagc 720  
 ctcagagatc tgcaagaact tgaagagcta catcaccaga tcccatttat cccttcagaa 780  
 gacagctggg cagtggccag tgagaagaat tctaacaagt atgtacagca agaaaagcag 840  
 aatacagcat ctcttagtaa agtaaatgcc agccgaattt taactaatga tctagagttt 900  
 gatagtgttt cagatcactc taaaacactt acaaatttctt ctttccaagc aaaacaagaa 960  
 agtgcattt cccagacata tcaatattgg gtacatttatt tggatcatga tagtttagca 1020  
 aataagtcaa tcacatataa aatgtttgaa aaaaccttaa gtggcacaaaa ttcaatttcc 1080  
 caagaaatta tgactctgt aaataatgaa gaattgacag atgaactatt aggttgcata 1140

gctgcagaat tattagctct ttagtggaaaa gataacaact cttgcacaaa aatggcaat 1200  
gaaacagatc ctgaaaacct aaatcttgc ctcagatgga gaggaaatc cccaaaagaa 1260  
atgggcagag agacaacaaa agtcaaataa cagaggata gtagtggct cagatataat 1320  
gacagggagg agaaatttct catctcaaataa gaaaagaaga tattttctga aaatagttt 1380  
aagtctgaag aacctatcct atggaccaag ggtgagattc ttggaaaggag agcttacggc 1440  
acagtataact gtggtctcac tagtcaagga cagctaataag ctgtaaaaca ggtggcttt 1500  
gatacctcta ataaatttgc tgctgaaaag gaataccgga aactacagga agaagtagat 1560  
ttgctcaaag cactgaaaca tgcacatttgc tggcctatt tgggacatg ctgcacagag 1620  
aacactgtga gcattttcat ggagttgtt cctggcttgc caatctctag tattataaac 1680  
cgtttggc cattgcctga gatgggttgc tgtaataataa cgaaacaaat acttcaagg 1740  
gttgcattatc tccatgagaa ctgtgtggta catcgacataa tcaaaggaaa taatgttgc 1800  
ctcatgccaa ctgaaataat aaagctgatt gactttggct gtgcaggcg tttggctgg 1860  
gcagggtttaa atggcacccca cagtgacatg cttaagtccaa tgcacatggac tccatattgg 1920  
atggcccccag aagtcatcaa tgagtctggc tatggacgga aatcagatataat ctggagcatt 1980  
ggttgtactg tggttggat ggctacaggg aagccctccac tggcttccat ggacaggatg 2040  
gcccgcattgt ttacatcgagacacccga gggctgatgc ctcccttacc agaccacttc 2100  
tcagaaaatg cagcagactt tggtgcgtatg tgctgtacca gggaccagca tgacgcaccc 2160  
tctgcctcc agtcctgaa gcactcccttgc ttggagagaa gtcactgaat atacatcaag 2220  
actttctcc cagttccact gcagatgtcc cctgttttgc ttgtggggaa tgatggctaa 2280  
gggatcttg tttccctact gaaaatttgc tctaaccggc tttaaggcaga tcctatggag 2340  
tcattaaactg aaagttgcag ttacatatttgc tgcacatggac ttattactca 2400  
tagtatacaga aaacatgttc ttaataacaa caaaaaacta tttcagtttgc tacagtttt 2460  
atgtccagg aactacatcc tctgtgttt tatatgacat ttcttttat tttggctgg 2520  
tcctgtcaat tttaatgttgc tttagttaaa ataaatttgc aaaaacaaaaaaa aaaaaaaa 2577

<210> 29  
<211> 2110  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7483334CB1

<400> 29  
ctagggtcgc cggggaaagcg gtttggaga gcccattggc actgcgtgag tggagcccg 60  
ctgtgtggat gccccagcat ggatgactac atgtcctgaa gaatgattgg ggagggtctcc 120  
ttcggcagag ctcttttgc ttagcttgc agcagtaatc agatgtttgc catgaaagaa 180  
ataaggcttc ccaagtctttt ctctaataaca cagaattctt ggaaggaggc tggcttttta 240  
gccaaaatga aacaccctaa tattgttgc ttcaaagaat catttgc tgaaggacac 300  
ttgtatatttgc ttagtggata ctgtgttgc gggatctaa tgcaaaagat taaacagcag 360  
aaaggaaatg tatttgc ttagtggata agacatgata cttaatttgc ttacccaaat gtgccttgg 420  
gtaaatcaca ttccacaagaa acgtgttgc tcaatgttgc tcaatgttgc 480  
ctcaactcaga atgaaaatg gaaattgggaa gactttggat ctggccgtct tctctccat 540  
ccgatggcat ttgttgc ttagtggaa actccttattt atgtgcctcc agaaatttgg 600  
gaaaacccgtc cttataacaa taaaatgttgc atctggctt tgggttgc tctgtatgaa 660  
ctctgttaccc ttaagcatcc atttgc ttagtggaa aatgttgc aaaaatctt cctcaagatg 720  
tgtcaagggttgc ttagtggaa ttttttttttgc ttttttttttgc ttttttttttgc 780  
aaggcgttgc ttagtggaa ttttttttttgc ttttttttttgc ttttttttttgc 840  
ggcgttgc ttagtggaa ttttttttttgc ttttttttttgc ttttttttttgc 900  
gaggaaatg ttagaagaaat aaaaatgttgc aagcataaca caccacaaaaaaaac 960  
cccacccatc ttagtggaaat gaagcataaca caccacaaaaaaaac 1020  
gatagaaagg gtagccatc tgatggaa agcattatg aaaaatgttgc tggaaatgttgc 1080  
ttgagaagag taaacagaga agaaaaatgttgc aataatgttgc tccatgttgc gaaagccatg 1140  
tcaccaaaatc ttcatagacg acagtgggag aaaaatgttgc ccaatgttgc tcttacagct 1200  
ttggaaaatg catccataact cacccatc ttaacagcag aggacgatag aggtggcttct 1260

gtaataaaagt acagcaaaaa tactactcgta aagcagtggc tcaaagagac ccctgacact 1320  
ttgttgaaca tccttaagaa tgctgatctc agcttggctt ttcaaatacata cacaatata 1380  
agaccagggtt cagaagggtt cttgaaaggc cccctgtctg aagaaacaga agcatcgac 1440  
agtgttgcgtt gagggtcacga ttctgtcatt ttggatccag agcgaacttga gcctggct 1500  
gatgaggagg acacggactt tgaggaggaa gatgacaacc cgcactgggt gtcagagct 1560  
aagaagcgag ctggatggca aggcctgtgc gacagataat gcctgaggaa atttcctga 1620  
gtcacgctga ggagagcctt cactcaggag ttcatgctga gatgatcatg agttcatgc 1680  
acgtatattt tcctttggaa acagaatgaa gcagaggaaa ctcttaatac ttaaaatcgt 1740  
tcttgattag ttcgttgagt ttgaaaagtc tagaactcct gtaagttttt gaactcaagg 1800  
gagaaggat agtggatgta gtgtgagcat cgggcttgc agtcccata aacagaaatg 1860  
ggatgctagc gtgccactac ctacttgcgtt gattgtggaa aattacttaa cctcttcaag 1920  
cccccaatttc ctcaaccata aaatgaagat aataatgcct acctcagagg gatgctgacc 1980  
acagaccttt atagcagccc gtatgatatt attcacatta tgatatgtt ttattattat 2040  
gtgactctt ttacatttcc taaaggttt agaattaaat atatthaattt atgatthaattt 2100  
aaaaaaaaaa 2110

<210> 30

<211> 7093

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7483337CB1

<400> 30

cgaggggacg cctcgcgacg gttcctggga gagctggcg cgcccttgc ctgcgcgc 60  
ttcgcgcgc cctcccccc cgccccctc aggattgagg aagtgcgtct gggccccc 120  
ccggcgcggg gggcagacgg cgggtggacg gccaggcccc gggccccc gttgtccgc 180  
ccggccccgc gtccccggagg agtcagctgt gtgtccagaa cgtgccatgg agacgcttaa 240  
cggtgcgggg gacacggggcg gcaagccgtc cacgcggggc ggtgaccctg cagcgcggc 300  
ccgcagacg gaaggcatcc ggcgcgcgt aaggcggggg gaccgcggcg ggcggggg 360  
cctgctggag gaggcctgcg accagtgcgc gtcccaagctg gaaaaggggcc agttctgag 420  
catccggca gcctatgggg atctggagat ggtccgcata ctactcagca agagactgg 480  
ggagctgccc accgagccca cggatgacaa cccagccgtg gtggcagcgt attttggaca 540  
cacggcagtt gtgcaaaaata cgctgcccac cgagccacg gatgacaacc cagccgtgg 600  
ggcagcgtat tttggacaca cggcagggtt gcaaggattt cttgagtcct taccagggtcc 660  
ctgcagtcctt cagcggctt tgaactggat gctggcctt gcttgcac gaggccac 720  
gggggtgtg aagctcctgg tcctgacgca cggggctgac cggagagct acgtgtcag 780  
gaagaatgag ttccctgtca tcgtgcctt gcccctgtat gcccctatca agtcaggaa 840  
tgaagacatt gcaatattcc tgcttcggca tggggcctat ttctgttcct acatcttgc 900  
ggatagtcct gacccagca aacatctgt gagaaggatc ttcatgaaag ccagtccctt 960  
gcccagcagt tatccggaa aaacagctt ccgtgtgaaa tggcccatt tcagactgcc 1020  
ctgggttagac cttagactggc tcataagacat ctcctgcctt atcacggagc tcgacccctt 1080  
tgccaaacttc ctggcgaccc tccctcggt tatccctgg ggcctcatca atctccggaa 1140  
gctgaaccc tccgacaacc acctggggg gtcgccttgc gtgcagtc gggacgaaat 1200  
catctgttcc aggctacttgg aaattgacat ttccagcaac aagttgtccc acctccctcc 1260  
tggattttt caccctctt aacttcaaaa actgacagct tcaaaaaattt gtttagaaaa 1320  
attgttcggaa gaagaaaatg ccaactactg gatagttt cggaaagctac aggaacttga 1380  
tatatctgac aataaaatttga cagaactccc tggccctgtt cttcactt tcaagtcctt 1440  
caattctctg aatgtctcca gaaacaaccc gaagggtttt ccagatccct gggcctgccc 1500  
tttgaatgt tgtaaagctt ccagaaatgc cctggaaatgt ctgcccagaca aaatggctgt 1560  
ctttggaaa aatcacctga aggtgttga ttctcagaa aacgcactca aagaagttcc 1620  
cctggactt ttccagctt atgcctcat gttcttgggg ttacaggggg accagctggc 1680  
ggcacttcca cctcaagaga agtggacctg caggcagctc aaaaccctgg atctctccag 1740  
aaaccaactt ggcaaaaatg aagatggact gaaaacgaag cgtattgcct tttcaccac 1800

cagaggctgc cagcgctccg ggactgaggc agagacaact atggagttca gtgcacatct 1860  
ggtaaccatt gtgttccctgt ctaacaactg taacctctgt gcatacacat gtgcagcaag 1920  
tgtgctggaa ttccggcct tcctaagtga gtcttggaa gtccttgc tgaacgacaa 1980  
ccacctcgac acagtcctc cctcggtttg cctactgaag agcttacatcg agctctactt 2040  
gggaaacaac cctggcctcc gggagctccc tcctgagctg gggcagctgg gcaacctctg 2100  
gcagctggac actgaagacc tgaccatcg caatgtgcct gcagaaatcc aaaaagaagg 2160  
ccccaaagca atgctgtctt acctgcgtgc tcagctgcgg aaagcggaaa agtgcacatc 2220  
gatgaagatg atcatcgtgg gtcccccgcg ccagggcaag tccaccctcc tggagatctt 2280  
acagacgggg agggccccc aggtggtgca tggagaggcc accatcgaga ccaccaagtg 2340  
ggagctccag agggcggctg gctcgagagc caaggtcaag gatggctgc gtgcagatc 2400  
cctgtgggtt gagtccgtgg agttcaacgt ctggacatc ggggacccgg ccagcatggc 2460  
caactgtcaac cagtgcctt tcacggacaa ggcctgtac gtgggtgtct ggaacctggc 2520  
gctggggag gaggccgtgg ccaacctcca ttctggctg ctcaacatcg agggcaaggc 2580  
cccaaacccc gtggtgcgg tggcgggac gcacccgttat ttaattgaag ccaagttccg 2640  
tgtggaaagg attgcaacgc tgcgtgccta tggctggca ctctgcgcct cccctccgg 2700  
ctccaggccc acaggcttcc cagacatcac ttccaaacac ttacatgaga ttccctgcaa 2760  
gagcctggaa ggtcaggaag ggctgcgaca gctgatttc cacgtcacgt gcagcatgaa 2820  
ggacgtccgc agcaccatcg gctgccagcg actggcaggg cggctgatcc ccaggagcta 2880  
cctgagccctg caggaggccg tgctggcaga gcagcagcgc cgcagccggg acgacgacgt 2940  
gcagtacctg acggacaggc agctggagca gctgggtggag cagacgcggc acaacgacat 3000  
caaggactac gaggacactgc agtcagccat cagcttcctc atagaaaccc gcaccctgt 3060  
ccatttcccg gacaccagcc acggcctgag gaaacctctac ttccctgcacc ctatttggct 3120  
ctccgaatgt ctgcagagga tcttaatata taagggtctt cggtcagttt ccaagaatgg 3180  
ggtgatcaga gcagaagacc tcaggatgt gctgggtgggg actggcttca cgcagcagac 3240  
ggaagagcag tacttccagt tccctggccaa gttttagatc gcccgcgg tcgccaatgaa 3300  
cagctaccc tcgccccatc tccttccatc taaacctggc ctggacaccc acgttatgcg 3360  
gcaccccaaca gccaacacca ttcagagggat attaagatg agttcgttc ccgttggctt 3420  
ctggcaaagg tttatagcac ggatgctgt cagccctggcg gagatggacc tgcagtttt 3480  
tgaaaacaag aagaatacta aaagcagaa cagaaagtc accatttaca gttttacagg 3540  
aaaccagaga aatcgctgtt gcacatttca agtaaaaga aatcagacca tctattggca 3600  
ggaagggctc ctggtcactt ttgatggggg ctacctcagt gtggaatctt ccgacgtgaa 3660  
ctggaaaaaaag aaaaaaagcg gaggaaatgaa aattgtttgc caatcagaag tgagggactt 3720  
ctcagccatg gcttcatca cggaccacgt caatttcctt attgatcagt gtttcccg 3780  
cctgacagcc acagagagcg acgggacgc actcatggag cagtcgtgc cctgcccgg 3840  
ctgcgagaca gcctggccc agcacacgg ccccaagttagaa attcagagg atgtcgtgaa 3900  
cttcgacatg gaagactgtg tcctgacggc catcgagccg gacttcatct cctgccccag 3960  
acacccggac ctccccgtgc cgctgcagga gctggccctt gaaactgttca tgaccgactt 4020  
cccgcccaagg ctttccctgg agaacagcaa gctggagcac agcaggacg agggcagcgt 4080  
cctggccag ggccgcgtg gcaccgtcat ctaccgggccc cggtaaccagg gcccgcctgt 4140  
ggccgtcaag cgcttcacca tcaaaaaattt caagaacttt gctaactgtac cggcagacac 4200  
catgctgagg cacctgcggg ccaccgtatc catagaagaaat ttctccgagt tccggcagga 4260  
ggccagcatg ctgcacgcgc tgcaaccc ctgcacatcg ggcgtcatcg gcatcagcat 4320  
ccacccgctc tgcttcgccc tggagctgc gcccgtcagc agcctcaaca ccgtgctgtc 4380  
cgagaacgcc agagattttt cctttataacc cctggagcac atgctcaccc aaaaaatagc 4440  
ctaccagatc gcctcgccccc tggcctaccc gcacaagaaa aacatcatct tctgtgaccc 4500  
gaagtcggac aacattctgg tgggtccct tgacgtcaag gagcacatca acatcaagct 4560  
atctgactac gggatttcga ggcagtcatc ccatgaggcc gcccgtcagc tggagggcac 4620  
tcctggctac caggccccag agatcaggcc tcgcattgtt tatgtgaga agtagatata 4680  
gttctccat tggatggtgc tctacgatg gctgtcaggaa cagcgcctg cactggccca 4740  
ccaccagctc cagattgcca agaagctgtc caagggcatc cggccgggttcc tggggcagcc 4800  
ggaggaagtg cagttccggc gactgcaggc gctcatgtat gatgtcgtggg acactaagcc 4860  
agagaagcga ccgcgtggccc tgcgggtt gaggccagatg aaggaccga ctttgcac 4920  
cttcatgtat gaactgtgt gttggaaagca gacgccttc ttctcatccc agggccagg 4980  
gtacaccgtg gtgttttggg atggaaaaga ggagtccagg aactacacgg tggtaacac 5040  
agagaaggcc ctcatggagg tgcagaggat gtgcgtccctt gggatgaagg tgagctgcca 5100  
gctccaggtc cagagatccc tggacagcgc caccgagaat tcctacctgg tcttagcggg 5160

cctcgccgat gggcttgtgg ctgtgttcc cgtgggtcgaa ggcaccccaa aggacagctg 5220  
ctcctacctg tgctcacaca cagccaaacag gtccaagttc agcatcgaa atgaagacgc 5280  
acggcagaac ccctacccag tgaaggccat ggaggtggtc aacagcgct ctgaggtctg 5340  
gtacagcaat gggccgggccc tccttgcata cgactgtgc tccctggaga tctgcaggcg 5400  
gctggagccc tacatggccc cctccatggt tacgtcagtc gtgtgcagct ctgagggcag 5460  
aggggaggag gtctgtctggt gcctggatga caaggccaaac tccttggta tgtaccactc 5520  
caccacccatc cagctgtgtg cccggtaactt ctgcggggc cccagcccc tcagggacat 5580  
gtttccctgtg cggcccttgg acacggaaacc cccggcagcc agccacacgg ccaacccaaa 5640  
ggtgcctgag ggggactcca tcgcggacgt gacatcatg tacagtgagg agctgggcac 5700  
gcagatccctg atccaccagg aatcaactcactc tgactactgc tccatgttct cctactcctc 5760  
atccccaccc cgccaggctg ccaggtcccc ctcaagcctc cccagctccc cagcaagtcc 5820  
ttccagttgtg ccttcttcca ccgactgcga ggactcagac atgctacata cgcccggtgc 5880  
tgccctccgac aggtctgagc atgacctgac cccatggac ggggagacat tcagccagca 5940  
cctgcaggcc gtgaagatcc tcgcgtcag agacccattt tgggtcccc ggcgcgggtgg 6000  
agatgttatac gtcatggcc tggagaagga ttctggcgcc cagcggggcc gatgttgc 6060  
cgctttaaaa gcccgagagc tgactccca tgggtgtctg gtggatgtc ccgtgggtggc 6120  
aaaggacact gttgtgtca cctttaaaa tgaaaacaca gatgttgc tggccgtctg 6180  
gaggggctgg ggcgcggagg agttcgacat ttcttaccag tcctacgagg agctggccg 6240  
gctggaggct tgcaactcgca agagaaggta attcctgtgg aatgactgtc acacatcaga 6300  
gctggctggc ccggggctgc agcctgactc ctctgccatc ggcctctagt tctccaagga 6360  
cctagaagac agatggagtt ctccccctgaa ctccttgcgt ctaagaagtg ctgagaagtt 6420  
actcgctgg cgggttgcctc agggttctctt ggttctctgg agcagagttc tctgaataacc 6480  
ccatccccca actgctgttattt ttacagcccc aggaagaca gtggatgtcag gctggggagcg 6540  
gcctccctgt gcctccccca tcagtttgc tggcaggagg tgcaggatcc tggcttgagc 6600  
tgggtcaaac aaagcaggcc cgggccttcc tgccatcccc aggtctcaga tggattaca 6660  
ctagaggccc tccgctggga agcaacttgcgt gtagggcagg agggggctg tgaccctgc 6720  
cctttcccccg ccagagacact caggctctca gcacattcca caggctcctg agtccccgag 6780  
gcctggggca gcttgggcaaa gccaagatca gatgtctctg tggctggaa ggtctccgtg 6840  
tggaaagcc cttgggggat cccgggttag gatgttgc ccatccagag aatgaatgag 6900  
ttcctttaag tgccaccggc agcaagccca gaggcacaca ttctgagtc acccgcttag 6960  
cctttacatt cctctccacc gacaaaagga agggaaaact caatcagcag gacttcagaa 7020  
agggccttgtt gtttatactt ttgtcaagta aatttggacg cagctggaaa cacaggcctg 7080  
tttggcact ata 7093

<210> 31  
<211> 1800  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 6035509CB1

<400> 31  
gctgcagagt gctttacttt caacaagatg gatgtttgtc ctgtttcccc gcctgttagtg 60  
cagtacaca gtcttggctc actgtacactt ctgcctcctg ggttcaagtg attctcctgc 120  
ctcagcctcc ttagtagctg ggattacagg aaacatctgt atggattattt tcactataat 180  
cctatgatgc ttggacttga atcacttcca gatcccacag acacctggaa aattatagag 240  
accattggta aaggcaccta tggcaagtc tacaaggtaa ctaacaagag agatgggagc 300  
ctggctgcag tgaaaattct ggatccagtc agtgcataatgg atgaagaaat tgaggcagaa 360  
tacaacattt tgcaatcat cccaaatgtt taaaagttt tggatgttt 420  
tacaaggcgg atcactgtgtt agggggacag ctgtggctgg tcctggagct gtgtaatggg 480  
ggctcagtc ctgagcttgtt caaaggctta ctcagatgtg gccagcggtt ggatgaagca 540  
atgatctcat acatcttgc tggggcccttc ttggggcttc agcatttgc caacaaccga 600  
atcatccacc gtgtatgtt ggggataaac attcttctga caacagaagg aggagttaaag 660  
ctcggttactt ttgggtttc agctcaactc accagtagacac gtctgcggag aaacacatct 720

gttggcaccc cgttctggat ggcccctgag gtcattgcct gtgagcagca gtagactct 780  
tcctatgacg ctcgctgtga cgtctggcc ttggggatca cagctattga actggggat 840  
ggagaccctc ccctcttga catgcattct gtaaaaacac tcttaagat tccaagaaat 900  
cctccaccta ctttacttca tccagaaaaa tgggtgaag aattcaacca ctttatttca 960  
cagtgtctta ttaaggattt taaaaggcga cttccgtca cacatctct tgaccaccca 1020  
tttattaaag gagtacatgg aaaagtctg ttctgc当地 aacagctggc caaggttctc 1080  
caagaccaga agcatcaaaa tcctgttgc aaaaccaggc atgagaggat gcataccaga 1140  
agaccttatac atgttggaaa tgctgaaaaa tactgcctt aggtatgtt ggtcaaccta 1200  
gagggtctgg atgaggtact aaatatttag tagacaattc tcattgaaga catttggat 1260  
atgtgaatgg tctgaacttt ctgtttaga ccatgtcctc ctaaggctcat ttggaaaaattt 1320  
aattgttgtt gtagctatgg gatgaagttc agggaggattt cagttgtgt gactatgatc 1380  
ctgtgctgtg tttattttaga tagccccat tagatgtaaag agaaaaggat ttggatttt 1440  
gcaataaagc tctttatattt gtagccttaa ttagtggatata ttagtgc当地 aaatattttt 1500  
tttgataaaa ttgtataaaa tatttcaattt aacccttaag aagttgtttt ttcttc当地aa 1560  
gaaagagctt catttaggaa aatagtgaag ttaatatacg ttgaattcta aatttgaagt 1620  
ctgtgataat ccccatattaa aatatgc当地 ttaatagag ctgttaattt gactggacct 1680  
tttatgtctg agtctactc tggggatgt taccttcaat gtctaaatca ctaaagtgt 1740  
atacaaagtg tttaattctg tattttagcc acctaggat tttagtgc当地 gctttgagaa 1800

<210> 32

<211> 6347

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7373485CB1

<400> 32

ggaagcgaga agccgcatca accatgttaag cagcttcgct tcctgccca accgtccgc 60  
gcctgaggag cccaccgccc ctctcgaaaa ccgacttccg ggggctgagc cgttgaagcg 120  
gaggctgggg cggggggcag cggcgccgc cggggcagga ggccgc当地 actcatgg 180  
ccacagatga taagacgtcc ccaacactgg actctgtctaa tgatttgc当地 ctagtgc当地ta 240  
ctagtc当地tc tcatctcaca cactttaaac ctttgactcc tgatcaagat gagccccctt 300  
ttaaatcagc ttatagttct ttgttaatc tcttc当地ttt taacaaagag agagcagaag 360  
gaggccaggg agaacagcag ctttgc当地tga gaagttggac cagccctc当地 ctcccttc当地 420  
ggacacagtc tggtaggtca cccacaccc ataaaaagca gcttaatgag gaactccagc 480  
ggcgcttc当地 agcatttagac acaagaagga aagcagaacc taccttggg ggtcatgacc 540  
ctcgtagc当地 tggtagctt cgaaggctca gcacagtattt aaaacgc当地 aaggaaatca 600  
tggagggaa aagccaggat agtgc当地tga aacaatactg gatgccagat agccaatgt 660  
aagagtgc当地 tgactgtatg gagaatttta caacctttag ggc当地c当地 cattgccgac 720  
tatgtggca gatttctgc agtgc当地tgc gtaatcaaga aatcccttggaa aaattttagg 780  
gctatacagg agacctccga gcttgc当地tgc attttagaaa aatagccctt agttagtgc 840  
attccacaga cagtaatttctt atggggaaag acttgc当地tgc tcttc当地tgc tctgcttgc 900  
ctgtgtctgt gcttgc当地tca agtgc当地tgc gacacccctt ggggatgtgg aaagccagcc 960  
gtaacatattt ttttagggat gatttggcc ggcaaagttt gattcatcca gattccctca 1020  
atactccctt ttcaacaaga cttgtatctg tgcaagagga tgctggaaa tctccctgctc 1080  
gaaatagatc agccagcattt actaacctgt cactggatag atctggctt cctatggatc 1140  
cttcatatga gacatctgtc agtccccagg ctaaccgaac atatgttagg acagagacca 1200  
ctgaggatga acgcaaaattt cttctggaca gtgtgc当地tgc aaaagacccctt tggaaaaaaa 1260  
tctgccc当地tca cagcactgtggaa atggagtttcc aggtatcaccgc ctactggatg agaacgc当地tca 1320  
ccaaactgc当地tca tggtagaaag gaatttagtca actggctaat cc当地aaatggg catattgcca 1380  
caagggcaca agctatagca attggacaag caatggatgaa tggacgttgg ctggattgtg 1440  
tttagtcatca cgaccagctt ttcagagatg agtgc当地tgc gtagatgc当地tca ctgc当地agatg 1500  
cagaatttcc tggagacgc当地tca tctccctgaca gtgactc当地tgc gaaactccctg gaaggacact 1560  
ctgagccatc ctgtttaaaa gacataaagt ttgtatgc当地tgc tgacacacaa cagatagctg 1620

aagaaggta cgataatttgc gctaagtatt tgatttctga cactggagga caacagctct 1680  
caataagtga cgctttcatc aaagaatcct tatttaatcg ccgagtagag gaaaaatcca 1740  
aagagctgcc tttcacaccc ttgggctggc atcataacaa cctggagctc ctgagggagg 1800  
agaatgggaa gaaacaagcc atggagaggt tgcttcagc taatcataac cacatgtgg 1860  
caactactcca gcagttgctc catagtgact cactgtcatc atcttggagg gacatcatcg 1920  
tgtcatttgtt ctgccaggtt gttcagacag tccgacactga tgtcaagaac cagatgtatg 1980  
acatggatata cgctcagttt gtccacatca aaaaaatccc aggtggaaag aagtttgatt 2040  
ctgtgggttcaatggcctt gttgtacca agaacattgc acataaaaag atgaattctt 2100  
gtattaaaaa ccctaaaatt cttctgttga agtggccat tgagtatctc tacagagaag 2160  
aaactaagtt tacattgcatt gatcctattt tgcttcagga aaggaaattt ttgaagaatt 2220  
atgtccagcg aatagttgtat gttcgaccca ccttgggttct tggtgagaaa acagtgtctc 2280  
ggattgccc gacatgttta ttggAACatg gcattactt ggtcattaaat gtAAAGTCAC 2340  
aagtttggaa acgaatcagt cgaatgaccc aagtgatgtt agtgatgtca atggaccaggc 2400  
tgcttacgaa accacgcctg ggcacttgc acaaattttt tatgcagata ttccagttgc 2460  
ctaataaca aaccaagaca ctgatgtttt ttgaagggttgc tccacagcac cttaggctgt 2520  
caatcaagct aagaggaggc tctgattatg agtggctcg agttaaggag atcctaataat 2580  
ttatgatctg tggtgttat cattctcaac tagaaatatc ctttctcatg gatgaatttg 2640  
ctatgcctcc cacattaatg caaaaccctt cattccattt cctgatttgc ggacgaggc 2700  
atgagggggc tgccaagag cgtacggtg gaggttccat cccctgggat cctgacatcc 2760  
ctcctgagtc tctgcctgt gatgatagca gttgtcttgc atcgaggatt gtgtttgaga 2820  
agggtgagca gaaaaataaaa aatcttccgc aggtgttgc ctctgtgaag catcaagaac 2880  
atagcacaac agcttgcctcg ggggtctcc cttgtgtttt ctttgcaccc gtaccggaaat 2940  
cattgtgcc actccctgtg gatgaccaac aagatgtttt aggcagcgag ctggcagaga 3000  
gtttcagca aacagtgtg ctgcaggatc cccaaagcca gataagagcc tttagagacc 3060  
ctctacagga tgacacttgc ttatatgttgc ctgaggaatg caccctctt gaagataaaac 3120  
gaaagacttta ttcttggcc ttaagcagg aattaaaaaaga tgtgatcttc tgtatctccc 3180  
cagtaatcac attccgagaa ccctttctt taactgaaaaa gggatgaga tgctctaccc 3240  
gagattattt tgcaagcag gtttacttgc ctccctcttca aataaaagaa ttcaaaagaaa 3300  
tggagaacag gaggaaagaaa cagctgtca gggatcttgc tgacttcag ggcattgt 3360  
gaagtattca gccaagtctt attcaagtctt taccctcaca tgagctatg agcactagaa 3420  
ttgctgagca tctggcgat agccagatg tgggtagaat gctggccat tattcgacca 3480  
gaggagaaat aattcagccca aaaaatttgc acccttttgc tcattcaaag gatgcattca 3540  
gtacttcaag tgcaaatca ggaagaaaa acgagggtga tgaagagaga gggcttattc 3600  
tgagtgtatgc tggtgtgtca acaaagggttgc actgtcttgc tccctattaaat caccagagac 3660  
tttgtgtctt cttcagcagc tcttctgttgc agtccagcaat tgctcctatg gctgtgtca 3720  
gtccttggat tgtaacaatg gaattttatg gaaagaatgt tcttacattt ggaatattttt 3780  
tagagagata ctgtttcagg ctttcttattt agtgcaccaatg catgttctgttgc gataacccca 3840  
tggtacatca tattcggcgc tttgttcatg gccaaggctg tggcagata atcctgaagg 3900  
agttggatttcc tccagttaccc ggtatcagc atacaatttgc tacatattcc tgggttagaa 3960  
tctgcacca ggtacacca gttgttgc tttccaatgt gtcctggat atgtcattttt 4020  
caaaataccat tgaacttggg tttatggc accagtataatc tcgcagagcc aacgctgagc 4080  
cctgtggatca ctccatccat catgattatc accagtattt tccctataac cagatgggtgg 4140  
cgtctttcag ttattctccc attcggcttc ttgttgcatttgc tggatcttc cccaaaatatt 4200  
tcattaaatgc tcaggccccca taaaaggatgc cccttcttca ggtatgttgc gacttcttgc 4260  
aaaaaggatcc acaggtatattt gttgcatttgc atgaaagact tgcatcttgc aaaaactgtata 4320  
cattttagtaa aacaagagag gaaaaatgg aagatattttt tgccacagaaa gagatggaaag 4380  
aagggtgatgtt caagaacttgg attgagaatgt tgcaagcaatg gctcatgtcttccctgttag 4440  
ataccctca gcaactgtcag tcggcttttgc agtcaactcat tggccaaatggaa caaagtctct 4500  
gtgaagtgtt gcaagcttgg aataacaggat tgccaggatctt tttccaacag gaaaagggtt 4560  
gaaagagacc ttcatgttccctt ccaagttctg gaagacttgg gcaaggggaa gaaagcaaga 4620  
taagtgcgtt ggtatgtatctt ccacggatata tttctccagg acttcagaat ggagaaaaaag 4680  
aggatcgctt ctttacacactt ttgtccagcc agagctccac cagttctactt catctccat 4740  
tgcctacgccc acctgttgc atgtcttgc acgtcaggatgggg agggccccctt gagcttagata 4800  
cagccagcgat ttccgaagat gtgtttgttgc ggcattttgc gggatccaca gacagccaaag 4860  
tgaaggaaaaa gtcacccatgtt aagccatctt ttgtccaaatgtt gttccagga aataagctata 4920  
atccatttttcc atttccatccat gatccagata aacactactt aatgtatgttgc aatgtatgttgc 4980

tgcccatatgc agtctgcgag aaggaaccca gctccatcat tgctttgct ctcagttgt 5040  
aagaataccg aatgcctta gaggaattgt ctaaagcgac tcagtgaaac agtgcgcgaa 5100  
aagggcttcc aacaaatagt acttcagata gcagacccaa gagtagcagc cctatcgat 5160  
tacctgaaat gagtgagga cagacaatc gtacaacaga aacagaacca caaccaacca 5220  
aaaaggcttc tggaaatgctg tccttcttca gaggacagc agggaaaagc cccgatctct 5280  
cttcccagaa gagagagacc ttacgtggag cagatgtgc ttactaccag gttgggcaga 5340  
caggcaagga ggggaccggag aatcaaggcg ttgagcctca agatgaagta gatggaggag 5400  
atacgcaaaa gaagcaactc ataaatcctc atgtggaact tcaattttca gatgctaattg 5460  
ccaagttta ctgtcggttc tactatgcgg gagagttca taagatgcgt gaagtgtttc 5520  
tggacagcag tgaggaagat ttcatcgtt ccctctccca ctcatcaccc tggcaggccc 5580  
ggggaggccaa atcaggagct gccttctatg caactgagga tgatagattt atttgaagc 5640  
aaatgcctcg tctgaaatgc cagtccttcc tcgactttgc accacattac ttcaattata 5700  
ttacaaatgc tggtaaccaa aagaggccca cggcggttgc caaaattctt ggagtttaca 5760  
gaattggta taagaactct cagaacaaca ctgagaagaa gttagatctc cttgtcatgg 5820  
aaaatcttt ctacgggaga aagatggcac aggttttga ttgaaaggc tctcttagga 5880  
atcggaatgt aaaaactgc actggaaaag agagttgtg tgggtcctg ctagatgaaa 5940  
atctcctaaa gatggttcga gacaaccctc tatatattcg ttctcattcc aaagctgtgc 6000  
tgagaaccc gatccatgt gactccatt tcctttctag ccacccatt atagattatt 6060  
ctttgcttgt tggcgagat gatactagca atgagctgt agttggaaatt atagattata 6120  
ttcgaacatt tacatggac aaaaagctt agatggttgt gaaatcaaca ggaatttttag 6180  
gtggacaagg taaaatgcac acagtgggtt ctccggagtt gtacaggact aggtttgtg 6240  
aggcaatgga caagtatttc ctaatggtac cagaccactg gacaggcttg ggtctgaatt 6300  
gctgaaatca agacatattt gaaatggact gtgaggaaaa ggggaaac 6347

<210> 33

<211> 1876

<212> DNA

<213> *Homo sapiens*

<220>

<221> misc\_feature

<223> Incyte ID No: 5734965CB1

<400> 33

tggggttcgg cgccggctacg tgcagaatcc gtctagctaa aatgttaattt cagattggac 60  
aagtactgtg gaggaaactgc aatgtctgggt ggagaacaga aaccagagag gtactatgtg 120  
ggtgtggacg ttgaaacagg cagtgtccgt gcagctctgg tggaccagag tgggttcctg 180  
ttggctttg cagaccagcc aattaagaat tgggagcccc agttcaacca ccatgagcag 240  
tcctccgagg acatctgggc tgcgtcgtgt gttgtcaca a gaaaagtgt acaaggatt 300  
gatttaaacc aaattcgagg acttgggtt gatgccacgt gttctctgg tgggttggat 360  
aagcagttc accattacc agtcaaccag gaaggggatt cccatcgaaa cgtcatcatg 420  
tggctggacc atcgagcagt cagtcaagtt aacaggatca atgagaccaa gcacagtgtc 480  
ctccagtagc tcgggggggt gatgtctgt gaaatgcagg ccccgaaact tctgtggctg 540  
aaagagaact tgagagagat ttgctggat aaggcgggac atttcttga tctccggac 600  
ttcttatctgt ggaaggcaac aggtgtcaca gcacggtctc tctgtccct ggtgtgtaa 660  
tggacatatt cagcagagaa aggctgggac gacagtttct gaaaaatgt tgggttggaa 720  
gactttgtt cagataatta cagcaaata gggaaaccaag tgctacctcc tggagcttct 780  
cttggaaatg ggctcacacc agaggcagca agagaccttgc gccttctccc tggattgcg 840  
gtgcagctt cactcattga tgcccatgca ggaggactag gagtgattgg ggcagatgtg 900  
agagggcacg gcctcatctg tgagggcag ccagtgcacgt cacggctggc tgcacatctgt 960  
ggAACGTCTT CTTGTCACAT GGGGATCAGC AAAGACCCGA TTTTGTACCC AGGCGTCTGG 1020  
GGGCCTTATT TCTCAGCCAT GGTACCTGGG TTCTGGCTGA ATGAAGGTGG TCAGAGCGTT 1080  
ACTGGAAAAT TGATAGACCA CATGGTACAA GGCATGCTG CTTTCCAGA ACTACAAGTA 1140  
AAGGCCACAG CCAGATGCCA GAGTATATAT GCATATTGA ACAGTCACCT GGATCTGATT 1200  
AAGAAGGCTC AGCCTGTGGG TTTCCTTACT GTTGATTAC ATGTTGGCC AGATTCCAT 1260  
GGCAACCGGT CTCCCTTAGC AGATCTGACA CTAAGGGCA TGGTCACCGG ATTGAAACTG 1320

tctcaggacc ttgatgatct tgccattctc tacctggcca cagttcaagc cattgctttg 1380  
 gggactcgct tcattataga agccatggag gcagcaggc actcaatcg tactctttc 1440  
 ctatgtggag gcctcagcaa gaatccccctt ttttgcaaa tgcattgcggc cattactggc 1500  
 atgcctgtgg tcctgtcgca agaggtggag tccgttcttg tgggtgctgc tggctgggt 1560  
 gcctgtgcct caggggattt cgcttctgta caggaagcaa tggcaaaaat gagcaaagtt 1620  
 gggaaagttg tggcccgag actacaggat aaaaaaact atgataagaa ataccaagta 1680  
 ttcctgaagc tgggtgaaca ccagaaggag tatttggcga tcatgaatga tgactgaaca 1740  
 gggcttgcag gtgctgatgc cagaagctt cgtgcatttgcatt cattaaagac ttctgtcatt 1800  
 tgatccatgt tcaagaccct tgaggtatttgcatt tctgtattgt cttcaataa 1860  
 agaatacataa catgtg 1876

<210> 34  
 <211> 1487  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7473788CB1

<400> 34  
 atgaggagtg gcgccgaacg caggggcagc agcgcgcgg cgtccccggg ctcgcgcgc 60  
 cccggccgcg cgcgcggcgc cggctccgac gcgcctcg 120  
 ggcgcggccccc gggccgggat ctcggcgat gtccgcgc 180  
 tggagcaagt ggaagaagag gatggctcg tccatgtcg 240  
 gtgtttgacg acaaggagga cgtgaacttc gaccacttc 300  
 aagggcagct ttggcaaggt gtgcattgtg cagaagcggg 360  
 atgaagtaca tgaacaagca gcagtgcac 420  
 gagctggaga tcctgcagga gatcgac 480  
 caggacgagg aggacatgtt catggctgtg gacctgtac 540  
 cacctgcacg agaacgtgca gttctcgag gacacgggtg 600  
 gcactggctc tggactaccc ggcgcggcc 660  
 aacattctcc tggatgagag aggacatgca cacctgaccg 720  
 atcaaggacg gggagcgggc gacggcatta gcaggcacca 780  
 atcttccact cttttgtcaa cggcgccgacc 840  
 gtgggggtga tggcttatga gctgctgcg 900  
 aacgcgtgg agtccctggt gcagctgtt 960  
 tggtccaagg agatggtggc cttgtcg 1020  
 ctctccagcc tccaggacgt gcaggcagcc cccggcgctgg 1080  
 ctgagcaga agagggtgga gcccggcttc gtgcacaaca 1140  
 cccaccccttgcg 1200  
 cgcctggcca agaacaagtc cccggacaaac 1260  
 tatcttcaag actgcctcg 1320  
 ctgaagagga gccaggaccc 1380  
 cggagcctgt ggaggacgag gcggacgctc 1440  
 tcggccggga gcccggcttgc 1487

<210> 35  
 <211> 1884  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3107989CB1

&lt;400&gt; 35

gaggtgacca atttctctc caaaagagaa aggaagttga taaaaaaaaag aatccatgct 60  
 ccaaaggcc agccaaatcc atctatggcc cccatgcac cacccagaaa gggttccag 120  
 actctcctgc aaaaggccaa ctctacttcc cggctccac ttcccctct tcgccacagg 180  
 agggtggcga aggattata acccacctct ttcttcagt tgccatggag acaagcccc 240  
 gtcccttcat tccttcttgt acctctctc ccaacgcagg cggaaaggag gcgccttagc 300  
 ccaaacatgc tgggggaggg gctggcgcc tcgacggcag ctgcggaaact aggccgaggg 360  
 acaaaggcta agttttcca tggtttggac tggatatcg tggaaactctg gtcaagctgg 420  
 tatattttga acccaaagac atcaactctg aagaagaaga ggaagaagtg gaaagtctta 480  
 aaagcatcg gaagtacctg acctccaatg tggcttatgg gtctacaggc attcgggacg 540  
 tgcacctcga gctgaaggac ctgactctgt gtggacgcaa aggcaatctg cactttatac 600  
 gctttccac tcatgacatg cctgcttttca ttcaaattggg cagagataaa aacttctcga 660  
 gtctccacac tgtcttttgt gccactggag gtggagcgt acaaattttag caggattttc 720  
 tcacaatagg tgatcttcag ctttgcaaac tggatgaact agattgtttt atcaaaggaa 780  
 ttttatacat tgactcagtc ggattcaatg gacggtcaca gtgctattac tttgaaaacc 840  
 ctgctgattc tgaaaagtgt cagaagttac catttattt gaaaaatccg tattctctgc 900  
 ttctggtaa cattggctca ggggttagca tcttagcagt atattccaaa gataattaca 960  
 aacgggtcac aggtactagt ctggaggag gaacttttt tggctctgc tggcttttta 1020  
 ctggctgtac cactttgaa gaagctcttga aatggcatc tcgtggagat agcaccaaaag 1080  
 tggataaact agtacgagat atttatggag gggactatga gaggtttggc ctgccaggct 1140  
 gggctgtggc ttcaagcttt ggaaacatga tgagcaagga aaagcgagat tccatcagca 1200  
 aggaagaccc cggccggcc acattggcata ccatcaccaa caacattggc tccattgctc 1260  
 ggatgtgtgc gttgaatgag aacatagaca gagttgtgtt tggtaat tttctcagaa 1320  
 tcaatatggt ctccatgaag ctgctggcat atggcatgaa tttttgtcc aaaggacaac 1380  
 tgaaagctct gttttggaa catgagggtt atttggagc cgttgggca ctgttggaaac 1440  
 tggtaat gactgatgac aagttagagac gagcagtggc ggaaacagcc tccaaaagg 1500  
 acagagaact aaaaattgc tgctggagaa ggtgaaagtc gcttgggac ggaagccaag 1560  
 ccattatggc agatgaaccc gctggattt gaaataattt aaaatccctc cagatgatct 1620  
 tttactcttta gttttgagc taatgatca aaacggggaa atataaaagg tttttttct 1680  
 gtataactgtt ttttttaaa aaaatggtgc agcgtggcca aacctaccaa ttgtatgcat 1740  
 taactttgaa aagttgtttg atgttaaga aggacctgat atgtaaagcgc tggcatttt 1800  
 tcttctgggg ttactgatc agtgtggta ttttaacttc atttagtaat tactcttagga 1860  
 gattttaccc tgacttataat ttcc 1884

&lt;210&gt; 36

&lt;211&gt; 1070

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7482887CB1

&lt;400&gt; 36

gcaaatcaca cagcatggca gctcccagtc ctccgcctc ttctgcattc cagacctgct 60  
 ctttaaaaac ctgggcattc cttccacaaa ttgaagagtg gaattttttt tcacctgctc 120  
 ttccctctgc tggcacagat cataaagtct tgctctttt ctatcacatc tcattattat 180  
 tttggctctt ttctacaagc aaggagcagc agggccctttt acattaccat tagtgaaggc 240  
 acttgagttt aatccgcaca acgaatctt ctttgcctg taatcccagc actttggaaa 300  
 gccaaggccggtt gttggatcacc tgaggtcagg agttcgagac cagcctggcc aatgtggtaa 360  
 aacctatctc tactaaaaat acaaaaaattt ggcaggtgtt ggtgggtggg gcctgtatc 420  
 tcagctactt gggaggctaa ggcaggagaa ttgttgaat ctggagaca gagttgcgg 480  
 tggcccaaga tagccactt gcactccagc ctttagcaaca agagcacaac tccatctcaa 540  
 aataataataataatttctt ggctccaagt ctcagctccc gcaccacctg acactgtcag 600  
 atccctcaggc catggccaaacttggagagca tcattatcaa tccgagtgtt gttcagcaca 660  
 gcctgggg tggaaatcatc aaatactctg agcagaaggatttctacccgttgcaccatga 720

agttccttcg ggccctctgag aaaccctgaa agccgcacta cactaacctg aaagaccacc 780  
cattcttccc ggaccttgtg aagtacatga actcaggca ggttgtgcc atggcctgg 840  
aggggctgaa tggcggcaaa acagggtaa ggatgctgg ggagaccaat tcattggct 900  
ctatgctaga gactattatt cgcaggact tctcgctaa aataggcggg aacgtcattg 960  
gtggcagtga ttcatcaca agtgctgaaa aagaaatcag cctatggtt aagcccaaag 1020  
aaccaggatga ctacagatct tggcttatg actgggtcta tgcatgata 1070

<210> 37

<211> 2890

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2963414CB1

<400> 37

gtgacccttc cctccccagg ccacggcagc cggccctcc cggcagacc tccgcacca 60  
ggctctggtaa acacagcaaa tgctccacgc tggacgggc cattgcctga tgccctgtaca 120  
tggtggcac ttagagacaa gattcctggg ccctgccttc catacactcc ccacgatctc 180  
ggaggaagct ctgaggacc cgcgtgaaac ccacagacag gaggacaact gcgcgtatgac 240  
agcaataaaag gccaagaagg agaaaggta ggaccgctga cagcccggt tgctgttggg 300  
agctgcctt tctacttcaa accttcctct agcagactgt gcaggacccc cccaccacca 360  
ccatctgccc ccatggttgt gcaaaaacagc gcaagacgccc gggacatgag ggcaaggcgtg 420  
cagctggagc cttccctgca ccaggtcggg gggcacatga gcgtgatgaa gtatgacgag 480  
catacgtgt gcaagccct cgtctcccg gaggcagaggt tctatgaaatc cctgcccgtg 540  
gccatgaagc gtttaccccc acagtacaaa ggttaccgtca cagtgacact ctgaaagac 600  
agcacaggcc atctcagctt gtttgcaccc ccagtgaagg agagccagga gcccctcaag 660  
gtctccacag agtcggccgc ggtggccata tggcagacgc tccagcagac caccggcagc 720  
aatggcagcg actgcacccct tgcccagtgg ccgcattgc agctggcacg ctcaaccaag 780  
gagagcccg ccaaggctct tctgagggtcc gagccccacc tcaacactcc agccttctcg 840  
ctgggtgaaag acaccaacgg aaaccagggtt gagaggaaga gcttcaaccc gtggggcctg 900  
caatgcacc accggccaccc gacccgcctg tgctccgagt acccagagaa caagcggcat 960  
cggttcttgt tgctggaaaa tggtagtgc cagtgacacgc atccctgtgt cctggatctg 1020  
aagatgggaa cccggcagca cggcgatgtat gcatcgaggaga agaagaaggc cccgcacatg 1080  
aggaagtgtg cgcagagcac ctcagctgc ctgggtgtgc gcatctgcgg catgcagggtt 1140  
tatcaaacag ataagaagta ctttctctgc aaagacaagt actatgaaag aaaactctca 1200  
gtggaggggt tcagacaagc cctctatcag ttccatcata atggaaagcca cttccggagg 1260  
gagctctgg agcccatcct gcaccagctc cggccctcc tctccatcat taggagccag 1320  
agttcatacc gtttctattc cagctcttc cttgtcatct atgatggca ggaaccacca 1380  
gaaagagccc caggcagccc gcatcctcac gaggctcccc aggcagccca cggtagctct 1440  
cccggtggtc tcaccaaggt tgacatccgc atgattgact ttgctcatac cacatacaag 1500  
ggctactgga atgagcacac cacctacgtt ggaccagacc ctggctatat ttttggcctg 1560  
gaaaaccta tcaggatcct gcaggatatc caagaggagg aatgaaactt cctgggctta 1620  
tctggattct tctggctat agatctaaa tagagacctg ttgggtgcta gggtagtcca 1680  
gacaccctt agatgtctt ataatagtcc tatctacctt caaaaaccat ctctatata 1740  
ggcagactat attaacagct gctgaacaaa tcagctctgg aggtgattcc acatccccctg 1800  
gcattatgtt ctaatgctgc tcattcgaga acagacagcc aggataaaatg ggcacccct 1860  
ggagtacact ggagggggca gccaagttt gaggccagca ttgctgacat tctggatata 1920  
ttgcattctaa aaatgtttac tcgttgcatt gctgcagttcc gcacaagctg tgaggcagaa 1980  
aacttgactt gaagcagcct tgaagagtga gttcatgagc tcattggttt tctcccttgc 2040  
tggactgctc gctccaaggc caggcagagc tcattgaatgc ctcttatctt cctaagcgg 2100  
gttttagtg acacaggatg aagcagaaga gatctaccca tctcacctgc tctgcacccca 2160  
gcttctaagt ggacaaagcc aagcccgaggc atgagctctg gcaaagcaag accccagatt 2220  
ctccatctt gctgtggaa aggagggtcc cttaacagcc ttttttcc ttttttccc 2280  
ccaaaatetc taaaatgag gaatcttta gcaactttt gatccccc ttctgcccaca 2340

ttctgaccat gagacgcggc ttgcagtggg ggtgaacgca cataaaaagg gaccactgac 2400  
 gtcctgcct actctctgct ttctatttat ttatttggg ggtgggtgg ggagtcagaa 2460  
 gaacctggag gacggaggaa accaggggca atgtttacaa gactggtgga caagtgtaaa 2520  
 tatggataaa gaacaaacag ttctaattaa ttccttcttc tgcagtaacgg aaaccttatta 2580  
 caatgccctt gagtcaagca ctgagatacg ttacccaatt agggaaataa atttgttaat 2640  
 aaaattgctg aggtcaccag tgattattgg tgcgttat tacccttcc atttgtttat 2700  
 tctgatcaca ctgtgtggta gttccaattt atgagcgact agcatatacc acaagaacag 2760  
 ttcaactgatt tcctacaatc cttcagggaa ctcgggtggaa aatgggtggct aataaaatat 2820  
 ttgcatgtat ctgcaaggga ggcaccagac ctgagaagtg gtcctttat ttgaatctca 2880  
 tacaatgtac 2890

<210> 38

<211> 5198

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7477139CB1

<400> 38

cgacacggag cacccttcta gcttcttcgt ctccaggact gacgctcagg ctccctcttc 60  
 gccttagccc aacttgcctt cccgcctcgc aaactccggt ttccctccac tcccaactct 120  
 tttcaactaca cgtttccctt cctctatctc ccacgcccacg aaccccgatc cccagactcc 180  
 tctctccgc cctcctcctt cctctctctt cccttcaact cttcatccgc ttccacctca 240  
 gactctgcgc gcacccaatt cagtcgccc ctcccgttcg gtcctcggaa gccatggcgg 300  
 gacctggggg ctggagggac agggaggta cggatctggg ccacctgccc gatccaactg 360  
 gaatatttctc actagataaa accattggcc ttgttactta tggcagaatc tatttggac 420  
 ttcatgagaa gactggtgca tttacagctg ttaaagtgtat gaacgctcgt aagacccctt 480  
 tacctgaat aggaaggcga gtgagagtga ataaatatca aaaatctgtt gggggagat 540  
 acagtgtatga ggaagaggat ctcaggactg aactcaacct tctgaggaag tactcttcc 600  
 acaaaaacat tggcatttc tatggagcat tttcaagct gaggccccctt ggtcagcggc 660  
 accaactttg gatggtgatg gagttatgtc cagcaggttc ggtcaactgtat gtagtgagaa 720  
 tgaccagtaa tcagagtttta aaagaagatt ggattgttta tatctgccga gaaatccctc 780  
 agggcttagc tcaccccttac gcacaccgg taatttcaccg ggacatcaaa ggtcagaatg 840  
 tgctgctgac tcataatgtc gaagtaaaac tgggtgat tggagtgtat gcccagggtga 900  
 gcagaactaa tggaaaagg aatagttca ttggacacc atactggatg gcacccgtagg 960  
 tgattgactg tgatgaggac ccaagacgtc cctatgatta cagaagtgtat gtgtggctg 1020  
 tgggaattac tggcattgaa atggctgaag gagccccctt tctgtgtaaac ttcaaccctt 1080  
 tggaaagctt cttcgattt ttgcggaaat ctgctccac agtcaatcc agccgatgg 1140  
 cccgtaagtt ccacaatttc atggaaaatgt gtacgataaa aaatttccgt tttcgcccta 1200  
 cttctgcaaa catgcttcaa caccatgtt ttccggatataaaaaatgaa cgacatgtt 1260  
 ttgagtctt aacaaggcat ttactggaa tcattaaaaa aagacagaaa aaaggaatac 1320  
 ctttgatctt tggaaagagaa gaagcttata aggaacagta caccgtgaga agattcagag 1380  
 gacccttgc cactcacgag cttctgagat tgccaaccag cagcagatgc agaccactt 1440  
 gagtcctgca tggggaaacc tctcagccaa ggtggctacc tgatcgagaa gagccacagg 1500  
 tccaggcact tcagcagcta cagggagcag ccagggtatt catgccactg caggctctgg 1560  
 acagtgcacc taagcctcta aaggggcagg ctcaggcacc tcaacgacta caagggcag 1620  
 ctgggtgtt catgccacta caggctcagg tgaaggctaa agcctctaaa cctctacaaa 1680  
 tgcagattaa ggcacccctca cgactacgga gggcagccag ggtgctcatg ccactacagg 1740  
 cacaggttag ggcaccccttgc cttctgcaagg tacagtccca ggtatccaa aagcagcagg 1800  
 cccagaccca gacatcgaa ccacaagatt tggaccaggt accagagaa ttccagggtc 1860  
 aagatcgatg accccgaacaa caaaggcagg gccaggcccc tgaacaacag cagaggcaca 1920  
 accagggtgcc tgaacaagag ctggagcaga accaggcacc tgaacagcca gaggtacagg 1980  
 aacaggctgc cgaggctgca caggcagaga ctgaggcaga ggaacctgag tcattacgag 2040  
 taaaatgccc ggtatttctg cccctgttat cacaagatca ccatgtctg ttgccactac 2100

atttggatac tcaggtgctc attccagtag agggcaaac tgaaggatca cctcaggcac 2160  
aggctggac actagaaccc ccacaggca ttggctcagt tcaagcactg atagagggac 2220  
tatcaagaga ctggcttcgg gcaccaaact caaataactc aaagccactt ggtccgttgc 2280  
aaaccctgat ggaaaatctg tcatcaaata gttttactc acaaccagaa caggcacggg 2340  
agaaaaaaatc aaaagttct actctgaggc aagcactggc aaaaagacta tcaccaaaga 2400  
ggttcagggc aaagtcatca tggagacctg aaaagcttga actctcgat ttagaagccc 2460  
gcaggcaaag gcgccaacgc agatgggaag atatcttaa tcagcatgag gaagaattga 2520  
gacaagttga taaaaccagt tggcgtcagt ggggtccttc agaccagttt attgacaata 2580  
gttcactgg tatgaagac ctgaagaaat atctcaaagg aaaaacaaca tttcataatg 2640  
ttcaagttgt tatctacaga gcagtttaagg ggaatgatga ttttgcacca aggtctaccg 2700  
ttcctcagcg gtctcttttga aacaagctc agaagccat tgacatcaga caaaggagtt 2760  
cgccaaatcg tcaaattgg ctggcagcat caggttattc aaagcacaaca attttagcag 2820  
gcaaaacaca gagctactgt ttaacaattt atatttgcacca agtcaagaaa gaagaatttc 2880  
aagaaggaat gaatcaaaag tgcaggag cccaaatggg attaggacccat gaaggccatt 2940  
gtatggca atgggtgaa tcttcttctg aggaagaaag tcctgtgact ggaaggaggt 3000  
ctcagtcatc accacccat tctactattt atcagaagtt gctgggtgac atccatgttc 3060  
cagatggatt taaagtagga aaaatatcac cccctgtata cttgacaaac gaatgggtag 3120  
gctataatgc actctctgaa atcttccgaa atgattggtt aactccggca cctgtcattc 3180  
agccacctga agaggatggt gattatgtt aactctatga tgccagtgct gatactgatg 3240  
gtgatgatga tgatgatgtt aatgataactt ttgaagatac ctatgatcat gccaatggca 3300  
atgatgactt ggataaccag gttgatcagg ctaatgatgt ttgtaaagac catgatgatg 3360  
acaacaataa gttgttgcattt gatgtaaata ataatttata tgaggccct agttgtccaa 3420  
gcttgcatttgc agggcaagct atggcagaga tggaaagctgc aagcaagatg gttatgatgg 3480  
aagtctggaa aagagggaaag cctacagagg ctatggaaagc catacagcca atagaagcca 3540  
tggaggaagt gcagccatgtt agggacaatg cagccattgg agatcaggaa gaacatgcag 3600  
ccaatatagg cagtgaaaga agggcagttt aggttgcattt aggttgcattt gtcgttcgaa 3660  
ccagtgaaga gagtggagcc cttggactca atggagaaga aaattgctca gagacagatg 3720  
gtccaggatt gaagagacccat gcgtctcagg actttgatata tctacaggag gagccagggtg 3780  
gtggaaatga ggcctcaaat gccatttgcattt caggtgctgc accgtcagca cctgtatcatg 3840  
agagtgcataa taaggacata tcagaatcat caacacaatc agattttctt gccaatcact 3900  
catctcatttcaaaatggatctg ctgtatgcataa ctttgcactt gccatcttcat 3960  
acgctggatt cgtagaagttt cctgaggaat caccaagca accctctgaa gtcaatgtt 4020  
acccactcta tgcatttgcattt gcatgtaaaa aaccactaat ccacatgtat gaaaaggagt 4080  
tcacttgcattt gatctgctgtt ggttcttgcattt gggagtttcaaa tttgtgttgcattt ggaacccat 4140  
ctaatttttgcattt tctgtatggac agaagtggaa aggtgcacat tactaaactt ataaggcgaa 4200  
gaccatccg ccagattcaat gtcatttgcattt cactcaattt gctgatttacc atctcaggc 4260  
ataagaacag acttcgggtt tatcatctga ctttgcatttgcattt gaaacaagatt ttgataatg 4320  
atccagaaag taaaagaagg caagaagaaa tgctgaagac agaggaagcc tgccaaagctt 4380  
ttgataagtt aacaggctgtt gacacttca gtccttcca acatgaagaa acaacatata 4440  
ttgcaatttgcattt ttttgcattt ggcacccaaag tcctttgtat 4500  
aaagcaactgc tattaaatgtt ttttgcattt ggcacccaaag tcctttgtat 4560  
ctattgttgcattt ttttgcattt ggcacccaaag tcctttgtat 4620  
tcgtatgcattt gtcatttgcattt ggcacccaaag tcctttgtat 4680  
tacctgattt gtcatttgcattt ggcacccaaag tcctttgtat 4740  
aagcaatgtt aacacttca aagaagatcc ttgaaatgtt gaaagacata ccatttcttca 4800  
tagctttgcattt gtcatttgcattt ggcacccaaag tcctttgtat 4860  
ctttgcattt gtcatttgcattt ggcacccaaag tcctttgtat 4920  
tcctgtgcac cccgggttgcattt ggcacccaaag tcctttgtat 4980  
gggtttactt catgacactt gggaaacttgcattt ggcacccaaag tcctttgtat 5040  
agtttccatgtt gtttatttgcattt ggcacccaaag tcctttgtat 5100  
atttcagaga taaaatgtt gtcatttgcattt ggcacccaaag tcctttgtat 5160  
acttttaatg tagcacagaa tagtttgcattt ggcacccaaag tcctttgtat 5198

<210> 39  
<211> 3969  
<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 55009053CB1

<400> 39

cttttttctt ttcagtgtgc ttcaaatgtc acgacacagg ttagctcagt cgacttgggg 60  
ctgctgagct ctggccctg ccagcctcac cgctcgacc ccccgatcc tccggactcc 120  
gctggctctg gccacgcgag gagcccacgc tagctccaaa gaatccccg agggcacgtg 180  
gaccgaggga gcccctgtga aggctcgga agactccgcg cgtcccgagc tcccgactc 240  
tgcagtggc cccgggtcca gggagccgct aagggtccct gaagctgtgg ccctagagcg 300  
gcggcgggag caggaagaaa aggaggacat ggagacccag gctgtggcaa cgcccccg 360  
tggccgatac ctcaagttt acatcgagat tggacgtggc tccttcaaga cggtgtatcg 420  
agggctagac accgacacca cagtggaggt ggcttgggtg gagctgcaga ctccgaaact 480  
gtcttagagct gagccgcgcg gcttctcaga ggaggtggag atgctcaagg ggctgcagca 540  
ccccaaacatc gtccgcttct atgattcgtg gaagtcgggtg ctgagggcc aggtttgcat 600  
cgtgctggc accgaactca tgacctcggt cacgctcaag acgtacctga ggccgttccg 660  
ggagatgaag ccgcgggtcc ttcaagcgtg gagccgccaa atcctgcggg gacttcattt 720  
cctacactcc cgggttcctc ccatcctgca ccgggatctc aagtgcgaca atgtctttat 780  
cacgggaccc tctggctctg taaaaatcg ggacctggc ctggccacgc tcaagcgcgc 840  
ctccttgc aagagtgtca tcgggacccc ggaattcatg gcccccgaga tgtacgagga 900  
aaagtacat gaggccgtgg acgtgtacgc gttcggcatg tgcgtctgg agatggccac 960  
ctctgagtagc ccgtactccg agtgcgagaa tgccgcgcaa atctaccgca aggtcacttc 1020  
ggcagaaaag ccgaacagct tccacaaggt gaagataccg gaggtgaagg agatcattga 1080  
aggctgcata cgcacggata agaacgagag gttcaccatc caggacctcc tggccacgc 1140  
cttcttccgc gaggagcgcg gtgtgcacgt ggaactagcg gaggaggacg acggcgagaa 1200  
gccgggcctc aagctctggc tgcgcattgg ggacgcgcgg cgccccggc gcccacggga 1260  
caaccaggcc atcgagttcc tggccatgc gggccggac gccccggcagg aggtggcaca 1320  
ggagatgtg gctctggct tggctgtga agccgattac cagccagttgg cccgtgcagt 1380  
acgtgaacgg gttgctgcca tccagcgaaa gcgtgagaag ctgcgtaaag caagggaaatt 1440  
ggaggcactc ccaccagagc caggacctcc accagcaact gtgcctatgg ccccccgtcc 1500  
ccccagtgtc ttccccctg agcctgagga gccagaggca gaccagcacc agcccttcct 1560  
tttccgcac gccagctact catctaccac ttccgattgc gagactgtatg gctacctcag 1620  
ctecccggc ttccctggatg cctcagaccc tgcccttcag cccctggg ggtgccatc 1680  
cagcctggct gagtccatc tctgcctgcc ctccggcttt gcccstatcca ttccacgttc 1740  
tggcccttgg agtgacttt ccccccggga cagctatgcc tcagatgcag ttccaggcct 1800  
tagcgatgtg ggagaaggga tgggacaaat gaggagaccc ccagggagga atctccggcg 1860  
cagaccccgta tcccgctgc gggtaactag tgcgtcgac cagaatgaca gagtggttga 1920  
gtgccagcta cagacccata acagcaagat ggtgaccttc cgatttgatc tggatgggaa 1980  
cagcccgaa gagattgcag ctgccatgtt atataacgag ttcatctgc ttccggagcg 2040  
agatggattt ctcaagacgga ttccggagat tattccagcga gtggagaccc tggtaagag 2100  
agacacttggc cccatggagg ctgctgaaga caccctaagc ccccaaggagg agccagcacc 2160  
attacctgccc ctggccgtcc ccctcccgaga cccatccaat gaagagctcc agagcagcac 2220  
ctccctggag cacaggagct ggacagcctt ctccacccctc tcatacttc tggactcc 2280  
tttgtctctt gaaaaacccat ttccccctgg aaccccccatt tccccaggtc ccattttccc 2340  
catcacttct ccccatgtc atcccagccc ctccccatcc tccccatttt ctccccaggt 2400  
ctccctcaat ccctctccac accccaccag ctctccactt ccattctcc tccagcacacc 2460  
cgagtttccg gtcccactct ctcaagtgtcc ctggagttct ctccccacga ctccctccacc 2520  
tacgttctct cccacttggc ttcaggtcac tcttagttcc cttttcttc tccgtgccc 2580  
ctccacttct tccttccct ccaccacagc agccctctc ctttctctgg cttagtgcctt 2640  
ctcaactggct gtgtatgactg tggcccgatc cctgctgtcc ccctcacctg ggctcccttc 2700  
ccagtcctct ccagcccccctc cttagtccctt cccttagcctg ccccttcccc ctcccggttc 2760  
tcctgggtggc cagggaaagcc ttccacccca cacagctgag gtggagagtg aggccctcacc 2820  
acctccctgtc cggcccccctcc caggggaagc caggctggcg cccatctctg aagagggaaa 2880  
gcggcagttt gttggggcgat tccaaatgtc ttcatccaag gaaccggctg agcccttc 2940

cttgcagcca acatccccca ctctctctgg ttctccaaaa ccttcaaccc ctcagctcac 3000  
ttcagagagc tcagatacag aggacagtgc tggaggcggg ccagagacca gggaaagctct 3060  
ggctgagagc gaccgtgcag ctgagggtct gggggcttga gttgaggagg aaggagatga 3120  
tgggaaggaa ccccaagttt ggggcagccc ccaacccctg agccatccca gcccagtgtg 3180  
gatgaactac tcctacagca gcctgtttt gagcagcggag gagtcagaaa gcagtggga 3240  
agatgaggag ttctgggctg agctgcagag tcttcggcag aagcacttgt cagaggttga 3300  
aacactacag acactacaga aaaaagaaaat tgaagatttgc tacagccggc tggggaaagca 3360  
gccccccaccc ggtattgtgg ccccaagctgc tatgtctgtcc agccgcaccc gcccctctc 3420  
caagggcagc ttccccaccc cccgcgcac cagcctacag cgctctgagc ccccaaggccc 3480  
tggcatcatg cgaagggact ctctgagtttgc cagcagcacc ggctcccaagg agcagcgggc 3540  
aagcaaggggg gtgacattcg cggggatgt tggcaggatgt tgaattcaga acagaagcca 3600  
tgtatctccc ccacaccagg gcccaccatg gagcttgttgc tctcagaatc tgatgcttc 3660  
tgatcaacaa aactgagcaa ggaagatccc aacactgaag gggtagaagg ccaggggggc 3720  
atggagagtg cagctccatt atagtgaaga gccaaacata tggtaactgt ttgtgtgt 3780  
gaggtgttag ttctgctgcc taccatcttc atctctagca cctccctgc caagagtcaa 3840  
ccactaagca atcccaccca agcctggatg ctcttagagg gcccactcc cagctgggg 3900  
agtgttagggg atatgctcac accacattag cagcaaccaa taaaaatgtct gaaacaaga 3960  
aaaaaaaaaa 3969

<210> 40

<211> 1803

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7474648CB1

<400> 40

atgggtgaaa gtggaaacca tcattttcag caaactaaca caggaacaga aaaccaaaca 60  
gcacatgttc tcactcataa gtgggagttt gacaatgaaa acatatggc acagggaggg 120  
gaacatcaca aactgggacc tgcgtatgggt tggaaaggcta ggagtggaa aacatttagga 180  
gaaataccta acgttaggcac actcacactc ctcaactggct atggggatgt ccagctgcca 240  
tgctgcaagg acactcaggc agcctatggaa gaaacccacg tggtgccggag tggaggcctt 300  
ctgccaacag ccagctgggactgaggcct gctgacagtc acacgggtac cagcgatgtat 360  
ccaggcgctc cgggtatccct gggggctgtc tccctgacca cgacccccc 420  
gtggggtttc ttcccgaggg tcccgccctt cgcagctgtct tttgataaaa gggcggagga 480  
acggggctgg ctgcttcccg agtccccagg tcccgcgagc ggcggcggtt tggcggtat 540  
gggggtgcggc gccagcagga aggtggtccc gggccacca gcgctggctt gggcaagca 600  
cgaaggtaa aaccaagccg gcgtcgaggagg cgccggccct gggcccgagg cggcggccca 660  
ggcggcgccag aggatacagg tggctcgctt ccgagccaag ttgcacccccc ggtccttgc 720  
cagtgcggcag tacaatttct tttgacatc tctgaacagg gagttcagag gatggaaaa 780  
aagagagcag gagcagcagc aaacaaggga aggaatttct attttcggag atatgacatc 840  
aaagctctta ttgggacagg cagtttcagc aggtttgtca gggtagagca gaagaccacc 900  
aagaaacctt ttgcaataaaa agtgtatggaa accagagaga gggaaaggtag agaagcgtgc 960  
gtgtctgagc tgagcgtcct gcgccgggtt agccatcggtt acattgtcca gctcatggag 1020  
atctttgaga ctgaggatca agtttacatg gtaatggagc tggctaccgg agggagctc 1080  
tttgatcgac tcattgctca gggatccctt acagagcggg atgcccgtcag gatcctccag 1140  
atgggttgcgt atgggatttag gtatggcat ggcgtcggag taactcatag gaatctaaag 1200  
cctgaaaacc tcttatacta tcattccaggta gaagagtcga aaattttat tacagatccc 1260  
ggtttggcat actccgggaa aaaaagtggt gactggacaa tgaagacact ctgtggacc 1320  
ccagagatca tagctccctga ggttttgcata aggaagcctt ataccagtgc agtggacatg 1380  
tggctcttgcgt gtgtgatcac atatgcttta cttagcgat tccctgcctt tgatgatgaa 1440  
agccagacaa ggctttacag gaagattctg aaaggcaaat ataattatac aggagaccc 1500  
tggccaagca ttcccactt ggcgaaggac ttatagaca aactactgtat tttggaggct 1560  
ggtcgtatcgca tgcgtatcgcc gccatccctt gggtagatcac catggctgca 1620

```
gggtcttcca tgaagaatct ccagagggcc atatccccaa acctcatgca gagggcctct 1680
ccccactctc agagtcctgg atctgcacag tcttctaagt cacattattc tcacaaaatcc 1740
aggcatatgt ggagcaagag aaacttaagg atagtagaaat cgccactgtc tgcgttttg 1800
taa                                         1803
```

<210> 41  
<211> 3472  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7483053CB1

<400> 41  
atggcgaagg cgacgtccgg tgccgcgggg ctgcgtctgc tgttgctgct gctgctgccg 60  
ctgctaggca aagtggcatt gggcctctac ttctcgaggg atgcttactg ggagaagctg 120  
tatgtggacc aggccggcgg cacggcccttgc ctgtacgtcc atgcccgtcg ggacgcccct 180  
gaggaggtgc ccagcttccg cctggccag catctctacg gcacgtaccg aacacggctg 240  
catgagaaca actggatctg catccaggag gacaccggcc tcctctaccc taaccggagc 300  
ctggaccata gctcctggga gaagctcagt gtccgcaccc gcccgttcc cctgctcacc 360  
gtctacccca aggtcttccct gtcacccaca tcccttcgtg agggcgagtg ccagtggcca 420  
ggctgtgccc gcgtatactt ctcccttc aacacccctt ttccagccctg cagctccctc 480  
aagccccggg agctctgctt cccagagaca agggcccttcc tccgcattcg ggagaaccga 540  
cccccaggca ccttccacca gttccgcctg ctgcctgtgc agttcttgcg ccccaacatc 600  
agcgtggcct acaggctctt ggagggtgag ggtctgcctt tccgctgcgc cccggacagc 660  
ctggaggtga gcacgcgctg ggccctggac cgcgagcagc gggagaagta cgagctggtg 720  
gccgtgtgca ccgtgcacgc cggcgcgcgc gaggaggtgg tgatggtgcc ctccccgggt 780  
accgtgtacg acgaggacga ctcggcgccc accttccccg cgggcgtcga caccgcccagc 840  
gccgtgggtgg agtcaagcg gaaggaggac accgtgggtgg ccacgctgcg tgtcttcgat 900  
gcagacgtgg tacctgcattt agggagctg gtgaggcggt acacaaggcac gctgtcccc 960  
ggggacaccc tggcccgagca gacccctccgg gtgaacact ggcaccaacga gacccgtggc 1020  
caggccaaacg gcagcttcgt gcgggcgacc gtacatgact ataggctggt tctcaaccgg 1080  
aacctctcca tctcggagaa ccgcaccatg cagctggcg tgctggtcaa tgactcagac 1140  
ttccaggggcc caggagcggg cgtcccttgc ctccacttca acgtgtcggt gtcggcggtc 1200  
agcctgcacc tgcccagtac ctactccctc tccgtgagca ggagggtcg ccgatttgcc 1260  
cagatcgaaa aagtctgtgt ggaaaactgc caggcgttca gtggcatcaa cgtccagtac 1320  
aagctgcatt cctctgggtgc caactgcagc acgcttagggg tggcacccctc agccgaggac 1380  
acctcgggga tcctgtttgtt gaatgacacc aaggccctgc ggcggcccaa gtgtgcccgg 1440  
cttcactaca tgggtgggtgc caccgaccag cagacctcta ggcaggccca ggcccagctg 1500  
cttgtaacag tggaggggtc atatgtggcc gaggaggcg gctgccccct gtctgtgca 1560  
gtcagcaaga gacggctggga gtgtgaggag tggggcggtc tgggctcccc aacaggcagg 1620  
tgtgagtggaa ggcaaggaga tggcaaagggtt atcaccaggaa acttctccac ctgctctccc 1680  
agcaccaaga cctgccccga cggccactgc gatgttggtt agacccaaaga catcaacatt 1740  
tgccctcagg actgcctccg gggcagcatt gttgggggac acgagcctgg ggagccccgg 1800  
gggattaaag ctggttatgg cacctgcacac tgcttccctg aggaggagaa gtgttctgc 1860  
gagccccgaag acatccaggaa tccactgtgc gacgagctgt gcccgcacggt gatcgccagcc 1920  
gctgtcccttct tctcccttcat cgtctcggtg ctgctgtctg cttctgcattt ccactgtac 1980  
cacaagtttgc cccacaagcc acccatctcc tcagctgaga tgacccctcg gaggccccgg 2040  
caggccttcc cggtcagcta ctcccttcc agtgcggcc ggcctcgct ggactccatg 2100  
gagaaccagg tctccgtggaa tgccttcaag atcctggagg atccaaagtgc ggaattccct 2160  
cggaagaact tggttcttgg aaaaactcta ggagaaggcg aatttggaaa agtggtaag 2220  
gcaacggcct tccatctgaa aggccagagca gggtacacca cggtgccgt gaagatgtctg 2280  
aaagagaacg cctcccccggag tgagcttcga gacccgtctgt cagagttcaa cgttctgtaaag 2340  
caggtcaacc acccacatgt catcaaatttgc tattggggccct gcagccaggaa tggccccgtc 2400  
ctccctcatcg tggagtacgc caaatacggc tccctgcggg gcttccctcg cggagccgc 2460

aaagtggggc ctggctaccc gggcagtgga ggcagccgca actccagctc cctggaccac 2520  
 ccggatgagc gggccctcac catggcgac ctcatctcat ttgcctggca gatctcacag 2580  
 gggatgcagt atctggccga gatgaagctc gttcatcgaa acttggcagc cagaaacatc 2640  
 ctggtagctg aggggcccga gatgaagatt tcggatttcg gcttgcctccg agatgtttat 2700  
 gaagaggatt cgtacgtgaa gaggagccag ggtcggattc cagttaaatg gatggcaatt 2760  
 gaatccctt ttgatcatat ctacaccacg caaagtgtatg tatggcttt tggtgtcctg 2820  
 ctgtgggaga tcgtgaccct agggggaaac ccctatcctg ggattccctcc tgacggcctc 2880  
 ttcaaccttc tgaagaccgg ccaccggatg gagaggccag acaactgcag cgaggagatg 2940  
 taccgcctga tgctgcaatg ctggaaagcag gagccggaca aaaggccggt gttgcggac 3000  
 atcagcaaag acctggagaa gatgatggtt aagaggagag actacttga cttgcggcg 3060  
 tccactccat ctgactccct gatttatgac gacggcctct cagaggagga gacaccgctg 3120  
 gtggactgta ataatgcccc cctccctcga gcctccctt ccacatggat tgaaaacaaa 3180  
 ctctatggca tgcagaccc gaactggcct ggagagatc ctgtaccact caccgagact 3240  
 gatggcacta acactgggtt tccaagatata ccaaattgata gtgtatatgc taactggatg 3300  
 ctttcaccct cagccggcaaa attaatggac acgtttgata gttaacattt ctttgaa 3360  
 ggtaatggac tcacaagggg aagaaacatg ctgagaatgg aaagtctacc ggcctttct 3420  
 ttgtgaacgt cacattggcc gagccgtt cagttccag gtggcagact cg 3472

<210> 42

<211> 1704

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7483117CB1

<400> 42

atggatgaca aagatattga caaagaacta aggcagaaat taaactttc ctattgttag 60  
 gagactgaga ttgaaggggca gaagaaagta gaagaaagca gggaggcttc gagccaaacc 120  
 ccagagaagg gtgaagtgc ggattcagag gcaaagggtt caccacccctt gactccctt 180  
 agcaacgtgc atgagctgca cacatctcg gaaaaagaca aagaaagtcc agatcagatt 240  
 ttgaggactc cagtgtcaca ccctctcaaa tgctctgaga caccagccca accagacagc 300  
 aggagaagc tgctgcccag tgacagcccc tctactccca aaaccatgct gagccggttg 360  
 gtgatttctc caacaggaa gcttccttcc agaggcccta agcatttga gctcacaccc 420  
 gctccctca agatgagat gacccattt gctctggtca atattaatcc cttcaactcca 480  
 gagtcctata aaaaattatt tcttcaatct ggtggcaaga ggaaaataag aggagatctt 540  
 gaggaagctg gtccagagga aggcaaggga gggctgcctg ccaagagatg tgtttacga 600  
 gaaaccaaca tggctcccg ctatgaaaaa gaattcttgg aggttggaaaa aattggggtt 660  
 ggcgaatttgcgtacatca caagtgcatt aagaggctgg atggatgtgt ttatgcaata 720  
 aagcgctcta tggaaacttt tacagaatta tcaaattgaga attcggctt gcatgaagtt 780  
 tatgctcacg cagtgttgg gcatcccccc catgtggtac gttactattc ctatggca 840  
 gaagatgacc acatgatcat tcagaatgaa tactgcaatg gtgggagttt gcaagctgct 900  
 atatctgaaa acactaagtc tggcaatcat tttgaagagc caaaactcaa ggacatccctt 960  
 ctacagattt cccttggcct taattacatc cacaactcta gcatggtaca cctggacatc 1020  
 aaaccttagta atatattcat ttgtcacaag atgcaaagtg aatcctctgg agtcatagaa 1080  
 gaagttggaaa atgaagctga ttggttctc tctgccaatg tgatgtataa aattgggtgac 1140  
 ctggccacg caacatcaat aaacaaaccc aaagtggaaag aaggagatag tcgcttcctg 1200  
 gctaattgaga ttttgcaga ggattaccgg caccttccca aagcagacat atttgcctt 1260  
 ggattaacaa ttgcagtggc tgcaggagca gagtcatgc ccaccaatgg tgctgcattgg 1320  
 caccatatcc gcaagggtaa ctttccggac gttcctcagg agctctcaga aagctttcc 1380  
 agtctgctca agaacatgat ccaacctgat gccgaacaga gaccccttcg agcagctctg 1440  
 gcccggaaa cagttctccg gccttcctg ggaaaaacag aagagctcca acagcagctg 1500  
 aatttggaaa agtcaagac tgccacactg gaaagggaaac tgagagaagc ccagcaggcc 1560  
 cagtcacccccc agggatatac ccatcatggt gacactgggg tctctggac ccacacagga 1620  
 tcaagaagca caaaacgcct ggtggagga aagagtgc当地 ggtcttcaag ctttacctca 1680

ggagagcgtg agcctctgca ttaa

1704

<210> 43

<211> 6298

<212> DNA

<213> *Homo sapiens*

<220>

<221> misc feature

<223> Incyte ID No: 7484498CB1

<400> 43

cgccggggcgg aacagatcgc agacctgggg gttcgcagag ccgcgcgtgg ggagatgttg 60  
aagtccaaat atggagcgcg gaatccctt gatgctggg ctgctgaacc cattgccagc 120  
cgggcctcca ggctgaatct gttcttccag gggaaaccac cctttatgac tcaacagcag 180  
atgtctccct tttcccgaga agggatatta gatgccctct ttgttctctt tgaagaatgc 240  
agtcagcctg ctctgatgaa gattaagcac gtagcaact ttgtccggaa gtattccgac 300  
accatagctg agttacagga gctccagcct tcggcaaaagg acttcgaagt cagaagtctt 360  
gtaggttggg gtcaactttgc tgaagtgcag gtggtaagag agaaaagcaac cggggacatc 420  
tatgctatga aagtgtgaa gaagaaggct ttattggccc aggagcagg ttcatttttt 480  
gaggaagagc ggaacatatt atctcgaaagc acaagcccggt ggatccccca attacagttat 540  
gcctttcagg acaaaaatca cctttatctg gtcatggaat atcagcctgg agggacttg 600  
ctgtcacttt tgaatagata tgaggaccag ttagatgaaa acctgataca gttttaccta 660  
gctgagctga tttggctgt tcacagcggtt catctgatgg gatacgtgca tcgagacatc 720  
aaggctgaga acattctcggt tgaccgcaca ggacacatca agctggggta ttttggatct 780  
gcccggaaaaa tgaattcaaa caagatgggta aatgccaaac tcccgattgg gaccccgat 840  
tacatggctc ctgaagtgc gactgtgatg aacggggatg gaaaaggcac ctacggccctg 900  
gactgtgact ggtggctcgt gggcgtgatt gcctatgaga tgatttattgg gagatcccc 960  
ttcgcagagg gaacctctgc cagaacccttc aataacatta tgaatttcca gcggtttttg 1020  
aaatttccag atgaccccaa agtgaggcagt gactttcttg atctgattca aagcttgg 1080  
tgcggccaga aagagagact gaagttgaa ggtctttgct gccatccctt ctctctaaa 1140  
attgactgga acaacattcg taactctcct ccccccttcg ttcccccacccct caagtctgac 1200  
gatgacacact ccaattttga tgaaccagag aagaattcgt gggtttcate ctctccgtgc 1260  
cagctgagcc cctcaggcctt ctcgggtgaa gaactgcccgt ttgtggggtt ttctgtacagc 1320  
aaggcactgg ggattcttgg tagatctgag tctgttggc cgggtctggta cttccctgccc 1380  
aagactagct ccatggaaaaa gaaactctc ataaaaagca aagagctaca agactctcag 1440  
gacaagtgtc acaagatgga gcagggaaatg acccggttac atcggagagt gtcagaggtg 1500  
gaggctgtgc ttagtcagaa ggagggtggag ctgaaggcct ctgagactca gagatccctc 1560  
ctggagcagg accttgcctac ctacatcaca gaatgcagta gcttaaagcg aagtttggag 1620  
caagcacgga tggaggtgtc ccaggaggat gacaaagcac tgcagcttct ccatgatata 1680  
agagagcaga gccggaaagct ccaagaaaatc aaagagcagg agtaccaggg tcaagtggaa 1740  
gaaatgaggt ttagtgcgaa tcagttgaa gaggatctt gtcagcaag aagacggagt 1800  
gatctctacg aatctgagct gagagagtct cggcttgcgt ctgaagaatt caagcggaaa 1860  
gcgcacagaat gtcaagcataa actgttgcgaa gctaaggatc aagggaagcc tgaagtgggaa 1920  
gaatatgcga aactggagaa gatcaatgct gaggcagcgc tcaaaattca ggagctccaa 1980  
gagaaactgg agaaggctgt aaaagccagc acggaggcga ccgagctgtc gcagaatatac 2040  
cgccaggccaa aggagcgcgcg cgagaggag ctggagaagc tgcagaaccg agaggattct 2100  
tctgaaggca tcagaaaagaa gctgggtggaa gctgaggaaac gccgcatttc tctggagaac 2160  
aaggtaaaaga gactagagac catggagcgt agagaaaaca gactgaagga tgacatccag 2220  
acaaaatccc aacagatcca gcagatggct gataaaaattc tggagctcga agagaaaacat 2280  
cgggaggccc aagtctcagc ccagcaccta gaagtgcacc tgaaacagaa agagcagcac 2340  
tatgaggaaa agattaaagt gttggacaat cagataaaaga aagacctggc tgacaaggag 2400  
acactggaga acatgatgca gagacacgag gaggaggccc atgagaaggg caaaattctc 2460  
agcgaacaga aggcgatgtat caatgctatg gattccaaga tcaagatccct ggaacagagg 2520  
attgtggAAC tgtctgaagc caataaactt gcagcaataa gcagttttt tacccaaagg 2580  
aacatgaagg cccaaagaaga gatgatttct gaactcaggc aacagaaaatt ttacctggag 2640

acacaggctg ggaagttgga ggcccagaac cgaaaactgg aggagcagct ggagaagatc 2700  
agccaccaag accacagtga caagaatcggtt gtcacatcgatggagacaatggatgg 2760  
gtcagtc tag agcacgagga gcagaaactg gagctcaagc gccagctcac agagctacag 2820  
ctctccctgc aggagcgcga gtcacagtgc acagccctgc aggctgcacg ggcggccctg 2880  
gagagccagc ttccgcaggc gaagacagag ctgaaagaga ccacagcaga agctgaagag 2940  
gagatccagg cactcacggc acatagagat gaaatccagc gcaaatttgc tgctttcg 3000  
aacagctgttgc tgtaatcac agacctggag gagcagctaa accagctgac cgaggacaac 3060  
gctgaactca acaaccaaaa cttctacttg tccaaacaac tcgatgaggc ttctggcgcc 3120  
aacgacgaga ttgtacacta gcaagtgaa gtggaccatc tccgcccggaa gatcacggaa 3180  
cgagagatgc agcttaccag ccagaagcaa acgatggagg ctctgaagac cacgtgcacc 3240  
atgctggagg aacaggtcat ggatttggag gcccctaaacg atgagctgttgc agaaaaagag 3300  
cggcagtggg aggccctggag gagcgtcctg ggtgtatgaga aatcccattt tgagtgtcg 3360  
gttcgagagc tgcaaggat gctggacacc gaaaaacaga gcagggcgag agccgatc 3420  
cgatcaccc agtctcgcca ggtgggtggag ctggcagtga aggagcaca ggctgagatt 3480  
ctcgctctgc agcaggctct caaagagcag aagctgaagg ccgagagcct ctctgacaag 3540  
ctcaatgacc tggagaagaa gcatgctatg ctggaaatgtc atgcccgaag cttacagcag 3600  
aagctggaga ctgaacgaga gctcaaacag aggttctgg aagagcagc caaattacag 3660  
cagcagatgg acctgcagaa aatcacatt ttccgtctga ctcaaggact gcaagaagct 3720  
ctagatcggg ctgatctact gaagacagaa agaagtgtact tggagtatca gctggaaaac 3780  
attcaggattc tctattctca tgaaaaggatgaaaatggaa gcaactatttca acaacaaacc 3840  
aaactcatttgc attttctgca agccaaaatg gaccaacctg ctaaaaagaa aaaggttcct 3900  
ctgcagatac atgagctgaa gctggccctg gagaaggaga aagctcgctg tgcaagacta 3960  
gaggaagccc ttcaagaagac ccgcacatcgatccctggccggagga agctgcccac 4020  
cgcaacgaa cggaccaccc acacccatcc acgcccagcca ccgcgaggca gcagatcgcc 4080  
atgtccgcca tcgtcggtc gccaigagcac cagcccagtgc catgagcct gctggccccc 4140  
ccatccagcc gcagaaagga gtcttcaact ccagaggaat ttagtccggc tcttaaggaa 4200  
cgcatgcacc acaatattcc tcacccgattc aacgttaggac tgaacatgcg agccacaaag 4260  
tgtgctgtgt gtctggatac cgtgcactt ggacgcccagg catccaaatg tctcgaatgt 4320  
caggtatgt gtcaccccaa gtgctccacg tgcttgcac ccacctgcgg ctgcctgt 4380  
gaatatgcca cacacttcac cgaggccctc tgccgtgaca aatgaactc cccaggtctc 4440  
cagaccaagg agcccgacag cagttgcac ctggaaagggtt ggatgaaggt gcccaggaaat 4500  
aacaacgag gacagcaagg ctgggacagg aagtagatggat tcctggaggg atcaaaagtc 4560  
ctcattatg acaatgaagc cagagaagct ggacagaggc cgggtggaaatg atttgagctg 4620  
tgcctcccg acggggatgt atctattcat ggtgccgtt gtcgttccga actcgcaaat 4680  
acagccaaag cagatgtccc atacatactg aagatggaaat ctcacccgca caccacgtc 4740  
tggcccgga gaacccctcta ctgtctagct cccagcttcc ctgacaacaca ggcgtgggtc 4800  
accgccttag aatcagttgt cgcagggtgg agatttctta gggaaaaagc agaagctgtat 4860  
gctaaactgc ttggaaactc cctgctgaaa ctggaaagggtt atgaccgtct agacatgaac 4920  
tgcacgctgc ctttcagtgc ccaggtgggtt tggtggca cccaggtgg gctctacgccc 4980  
ctgaatgtct tgaaaaactc cctaaccat gtcccaggaa ttggagcagt ctgcctaaatt 5040  
tatattatca aggacctgga gaagctactc atgatagcag gagaagagcg ggcactgtgt 5100  
cttgcgtggacg tgaagaaaatg gaaacagtcc ctggccctgg cccacctgc tgccaggccc 5160  
gacatctcac ccaacatccc tgaagctgtc aagggtgtcc acttgcgttgg ggcaggcaag 5220  
attgagaacg ggctctgtcat ctgtgcagcc atgcccagca aagtcgtcat tctccgtctac 5280  
aacgaaaacc tcagcaaata ctgcacccggaaatggatgaaatg agacctcaga gcccgtc 5340  
tgtatccact tcaccaatta cagtagtccat attggaaatc ataaattctt cggatgtc 5400  
atgaagcgtt acacgctcgaa ggaatttctg gataagaatg accattccat ggcacctgt 5460  
gtgtttcccg cctcttccaa cagctccctt gtctcaatcgatggatggatggatgg 5520  
cagcgagagg agtacttgcgtt gttttccac gatattggatggatggatggatggatgg 5580  
agacgttagcc gcacagacga tctcaagtgg agtgcgttac ctttggccctt tgcctacaga 5640  
gaaccctatc tggatgtgac ccacttcaac tcactcgaaatggatggatggatggatgg 5700  
tcctcagcag ggacccctgc ccgagcgtac ctggacatcc cggatgtc gggccctggc 5760  
cctgcctttt cctcaggagc gatattcttgcgtt accaggatggatggatggatggatgg 5820  
atttgctgtcat gggaaaacccctt cgtgaaggatggatggatggatggatggatggatgg 5880  
accccccgcgca gcagccccaa caagcgaggc ccacccacgt acaacgagca catcacaac 5940  
cgcgtggccctt ccagcccgaa gggccctggc accccggatggatggatggatggatgg 6000

ccccaccgct accgcgaggg gcggaccgag ctgcgcaggg acaagtctcc tggccgcccc 6060  
ctggagcgag agaagtcccc cggccggatg ctcagcacgc ggagagagcg gtccccggg 6120  
aggctttg aagacagcg cagggggccgg ctgcctgcgg gagccgtgag gaccccgctg 6180  
tcccaggtga acaaggtctg ggaccagtct tcagtataaa tctcagccag aaaaacccaac 6240  
tcctcatttt gatctgcagg aaaacaccaa acacactatg gaactctgct gatggga 6298

<210> 44

<211> 5454

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7638121CB1

<400> 44

cacgcacacc gcacgtacgg gttggggccc agctgggtta taagcgtat cccatgccc 60  
cctgcccagg ctggggggca tttgcacatc tgcaaaggcc tcccaagcctg tcccaagccct 120  
gccccagcct gggaccccca cattctactc accgtgtctc ctcagagggg ccagaaccct 180  
ccactggga gaggcaagtg gcggtgaact tgggttccat aggaccctgt ccctgagagc 240  
gacagcttag ttagttagct ccactggccc caccaactcc ttctgatcac ctggccagct 300  
gaggtcagag tgggagagggc agtgggttca ttgaaggagt actcctaact gtcagaagcc 360  
tggcggtca ggatggggtg ctgtcgcttgg ggtgttcaact tgcccacagt 420  
gtatctcagg gtctcaccaa ccatccaagc atgttaggt gtggctggca cccagggtt 480  
tgtggctggg gaggtggctt ccacagtcc ctcctgccc tcccaaggccc cccatccatg 540  
caggtaacca tcgaggatgt gcaggcacag acaggcggaa cggcccaatt cgaggctatc 600  
attgagggcg acccacagcc ctcggtgacc tggtaacaagg acagcgttca gctgggtggac 660  
agcacccggc ttagccagca gcaagaaggc accacatact ccctgggtct gaggcatatg 720  
gcctcgaagg atgcccggct ttacacctgc ctggcccaaa acactgggttgg ccaggtgctc 780  
tgcaaggccag agctgctggt gcttgggggg gacaatgagc cggactcaga gaagcaaagc 840  
caccggagga agtgcactc cttctatggat gtcaaggagg agattggaaagg gggcgtgtt 900  
ggcttcgtaa aaagagtgc gacaaaaggaa aacaagatct tgcgtctgc caagttcatc 960  
cccctacgga gcagaactcg ggcccaggca tacaggagc gagacatctt ggcggcgctg 1020  
agccacccgc tggcacggg gctgctggac cagtttgaga cccgcaagac cctcatcctc 1080  
atcctggagc tgcgtctatc cgaggagctg ctggaccggcc tgcgtacagaa gggcgtgtt 1140  
acggaggccg aggtcaaggt ctacatccag cagctgggtgg aggggctgca ctacactgcac 1200  
agccatggcg ttctccaccc ggacataaaag ccctctaaca tcctgatggt gcatcctgccc 1260  
cgggaaagaca ttaaaaatctg cgactttggc tttggccaga acatcaccccc agcagagctg 1320  
cagttcagcc agtacggctc ccctgagttc gtctcccccg agatcatcca gcagaaccct 1380  
gtgagcgaag cctccgacat ttggggccatg ggtgtcatct cctacctcag cctgacctgc 1440  
tcatccccat ttggccggcga gagtgaccgt gcccaccccttcc tgaacgtctt ggagggggcgc 1500  
gtgtcatgga gcagccccat ggctgcccac ctcagcgaag acgccaagaat cttcatcaag 1560  
gctacgctgc agagagcccc tcaggcccg cctagtgccgg cccagtgcct ctcccacccc 1620  
tggttcttga aatccatgcc tgcggagag gcccacttca tcaacaccaa gcagctcaag 1680  
ttccctctgg cccgaagtctg ctggcagctg tccctgtatg gctacaagtc catcctgggt 1740  
atgcgttcca tccctgagct gctgcggggc ccacccgaca gcccctccct cggcgttagcc 1800  
cgccacccctc gcagggacac tggtggtcttcc tccagttccct cctccctctc tgacaacgg 1860  
ctcgccccat ttggccgggc taagtctactg ccacccctccc cggtgacaca ctcaccactg 1920  
ctgcacccccc gggcttccct gcccggccctcg gccagcctgc ctgagggaaagc cgaggccagt 1980  
gagcgctcca cccgaggcccc agtcccgctt gcatctcccg aggggtgcgg gcccacccggcc 2040  
gcccagggtc gcgtgccccg gcacagcgctc atccgcagcc tggttctacca ccaggcgggt 2100  
gagagccctg agcacggggc cctggccccc gggagcagggc ggcacccggc cccggccgg 2160  
cacctgctga agggcgggtt cattgcgggg ggcgtgcccag gcctgcgcga gcccactgt 2220  
gagcaccggcg tgctggagga ggaggccggc agggaggagc aggccacccct cctggccaaa 2280  
gccccctcat tcgagactgc ctcggcttgc ctcggctctg gcacccactt ggcggccggc 2340  
cacagccact ccctggaaaca tgactctccg agcacccccc gcccctctc ggaggccctgc 2400

ggtagggcac agcgaactgcc tttagcccc tccgggggg cccatatcg ggacatgggg 2460  
caccctcagg gctccaagca gttccatcc actggtgcc acccaggcac tgctcagcca 2520  
gagaggccat cccggacag cccttgggg cagccagccc ctttctgca ccccaagcag 2580  
gttctgccc cccaggaggg ctgcagcccc caccagcag ttgcccattg ccctcctggc 2640  
tccttcctc caggatctt caaagaggcc cccttagtac cctcaagccc ttcttggga 2700  
cagccccagg cacccttgc ccctgccaaa gcaagcccc cattggactc taagatgggg 2760  
cctggagaca tctctttcc tgggaggcca aaacccggcc cctgcagttc cccagggtca 2820  
gcctcccagg cgagcttcc ccaagtgagc tccctcaggg tgggctcctc ccaggtggc 2880  
acagagcctg gcccctccct gnatgcggag ggctggaccc aggaggctga ggatctgtcc 2940  
gactccacac ccaccccttgc gcggcctcag gaacagggtga ccatgcgca gttctccctg 3000  
ggtagtcgcg ggggtacgc aggctggct ggctatggca ctttgcctt tggtggagat 3060  
gcagggggca tgctggggca gggcccatg tggccagga tagcctggc tgggtccctc 3120  
tcggaggagg aggagcagga ggaggccagg gctgagttcc agtcggagga gcagcaggag 3180  
gccaggctg agagccact gccccaggc agtgcaggc ctgtgcctga ggtcggcagg 3240  
gtccccacca ggagctctcc agagccacc ccatgggagg acatcggca ggtctccctg 3300  
gtgcagatcc gggacctgtc aggtgatgcg gagggggccg acacaatatc cttggacatt 3360  
tcggagggtgg accccgccta cctcaacctc tcagacctgt acgataatcaa gtaccccttca 3420  
ttcgagttt aatgtcttcag gaaagtcccc aagtccgctc agccagagcc gcctccccc 3480  
atggctgagg aggagctggc cgagttcccg gagggccacgt gggccctggcc aggtgaactg 3540  
ggccccacg caggccttgg aatcacagag gatcagagg atgtggacgc gctgctggca 3600  
gaggctgccc tggcaggaa ggcgaagtgg tcctcgccgt caccgcgcct cttccacttc 3660  
cctgggaggc acctgcccgt ggacgagcct gcagagctgg ggctgcgtga gagagtgaag 3720  
gcctccgtgg agcacatctc ccggatcctg aaggccaggc cggaaggctt ggagaaggag 3780  
ggccccccca ggaagaagcc aggccttgc tccttccggc tctcaggctt gaagagctgg 3840  
gaccgagcgc cgacattccct aaggagctc tcagatgaga ctgtggctt gggccagtca 3900  
gtgacactgg cctgcccagg gtcagcccag ccagctgcgc aggccacctg gagaaagac 3960  
ggagcccccc tggagagcag cagccgtgtc tcacatctctg ccaccctcaa gaacttccag 4020  
cttctgacca tcctgggtgt ggtggctgag gacctgggtg tgtaacacctg cagcgtgagc 4080  
aatgcgtgg ggacagtgc caccacggc gtccctccgg aaggcagagcg cccctcatct 4140  
tcgccccatcc cggatatcgg ggaggtgtac gggatgggg tgctgctggc tggaaagccc 4200  
gtggaaatcc acggccctgt gacctacatt gtgcagtgc gcttagaagg cggcagctgg 4260  
accacactgg cctccgacat ctttgcactc tgctacactg ccagcaagct cttccgggtt 4320  
ggcacctaca cttccgcac ggcatgtgtc agcaaggcag gaatgggtcc ctacagcagc 4380  
ccctcgagc aagtccctctt gggagggccc agccacctgg cctctgagga ggagagccag 4440  
gggcggtcag cccaaaccctt gcccagcaca aagaccttcg cattccagac acagatccag 4500  
agggccgct tcagcgtgtt gggcaatgc tgggagaagg ccagcggcg ggcgtggcc 4560  
gccaagatca tcccctacca ccccaaggac aagacagcag tgctgcgcga atacgaggcc 4620  
ctcaaggggcc tggccaccc gcacccggc cagctgcacg cagcttaccc cagccccgg 4680  
cacctgtgc tcacatccgg agtgtgttcc gggcccgagc tgctccctg cttggccgag 4740  
agggccctctt actcagaatc cgaggtgtac gactacctgt ggcagatgtt gatgtccacc 4800  
cagtacactgc acaaccagca catccgtcac ctggacactg ggtccgagaa catgatcatc 4860  
accgaataca acctgctcaa ggtcggtggc ctgggcaatg cacagagcct cagccaggag 4920  
aaggtgtgc cctcagacaa gttcaaggac taccttagaga ccatggctcc agagctcctg 4980  
gagggccagg ggctgttcc acagacagac atctggccca tcgggtgtac agccttcatc 5040  
atgctgagcg ccgagttaccc ggtgagcagc gaggggtgcac ggcacactgca gagaggactg 5100  
cgcaaggggc tggccggct gagccgtgc tacgcggggc tggccggggc cggccgtggcc 5160  
ttcctgcgc acaactctgtc cggccaggccc tggggccggc cctgcgcgtc cagctgcctg 5220  
cagtgcggcgt ggctaaacaga ggagggccc gcctgttgcg gggccggcc cgtgacccctc 5280  
cctaccgcgc ggctgcgcgtt tcctgcgcgc aatcgcgaga agagacgcgc gctgctgtac 5340  
aagaggcaca acctggccca ggtgcgtga ggtcgcccc ggcacacccc ttggtctccc 5400  
cgctgggggt cgctgcagac ggcacataa aaacgcacag cggccgaga aaaa 5454